

# GLow\_30\_1\_2\_distilbert

August 10, 2022

## 0.1 Install packages

```
[ ]: !git clone https://github.com/megagonlabs/ditto
      %cd ditto
      !pip install -r requirements.txt
      !git clone https://github.com/rit-git/Snippest_public
      %cd Snippest_public
      !pip install -r requirements.txt
      %cd ..

      import nltk
      nltk.download('stopwords')
```

```
Cloning into 'ditto'...
remote: Enumerating objects: 291, done.
remote: Total 291 (delta 0), reused 0 (delta 0), pack-reused 291
Receiving objects: 100% (291/291), 26.87 MiB | 8.21 MiB/s, done.
Resolving deltas: 100% (142/142), done.
Checking out files: 100% (128/128), done.
/content/ditto
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting gensim==3.8.1
  Downloading gensim-3.8.1-cp37-cp37m-manylinux1_x86_64.whl (24.2 MB)
    |                               | 24.2 MB 1.5 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
    |                               | 14.5 MB 51.8 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
    |                               | 689 kB 72.3 MB/s
Collecting scipy==1.3.2
  Downloading scipy-1.3.2-cp37-cp37m-manylinux1_x86_64.whl (25.2 MB)
    |                               | 25.2 MB 46.5 MB/s
Collecting sentencepiece==0.1.85
  Downloading sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
    |                               | 1.0 MB 66.1 MB/s
Requirement already satisfied: sklearn==0.0 in
```

```

/usr/local/lib/python3.7/dist-packages (from -r requirements.txt (line 6)) (0.0)
Collecting spacy==3.1
  Downloading
spacy-3.1.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.4 MB)
| 6.4 MB 59.0 MB/s
ERROR: Could not find a version that satisfies the requirement
torch==1.9.0+cu111 (from versions: 1.0.0, 1.0.1, 1.0.1.post2, 1.1.0, 1.2.0,
1.3.0, 1.3.1, 1.4.0, 1.5.0, 1.5.1, 1.6.0, 1.7.0, 1.7.1, 1.8.0, 1.8.1, 1.9.0,
1.9.1, 1.10.0, 1.10.1, 1.10.2, 1.11.0, 1.12.0, 1.12.1)
ERROR: No matching distribution found for torch==1.9.0+cu111
Cloning into 'Snippext_public'...
remote: Enumerating objects: 413, done.
remote: Counting objects: 100% (22/22), done.
remote: Compressing objects: 100% (8/8), done.
remote: Total 413 (delta 15), reused 14 (delta 14), pack-reused 391
Receiving objects: 100% (413/413), 21.06 MiB | 8.28 MiB/s, done.
Resolving deltas: 100% (196/196), done.
/content/ditto/Snippext_public
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting gensim==3.8.1
  Using cached gensim-3.8.1-cp37-cp37m-manylinux1_x86_64.whl (24.2 MB)
Collecting numpy==1.19.2
  Using cached numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
Collecting regex==2019.12.20
  Using cached regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
Collecting spacy==2.2.3
  Downloading spacy-2.2.3-cp37-cp37m-manylinux1_x86_64.whl (10.4 MB)
| 10.4 MB 29.6 MB/s
Collecting sentencepiece==0.1.85
  Using cached sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
Requirement already satisfied: sklearn==0.0 in /usr/local/lib/python3.7/dist-
packages (from -r requirements.txt (line 6)) (0.0)
Collecting tensorboardX==2.0
  Downloading tensorboardX-2.0-py2.py3-none-any.whl (195 kB)
| 195 kB 70.6 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
| 753.4 MB 7.2 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
| 56 kB 5.6 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
| 884 kB 61.3 MB/s
Collecting jsonlines==1.2.0

```

```

    Downloading jsonlines-1.2.0-py2.py3-none-any.whl (7.6 kB)
Collecting nltk==3.4.5
    Downloading nltk-3.4.5.zip (1.5 MB)
      |                               | 1.5 MB 62.4 MB/s
Requirement already satisfied: scipy>=0.18.1 in
/usr/local/lib/python3.7/dist-packages (from gensim==3.8.1->-r requirements.txt
(line 1)) (1.7.3)
Requirement already satisfied: smart-open>=1.8.1 in
/usr/local/lib/python3.7/dist-packages (from gensim==3.8.1->-r requirements.txt
(line 1)) (5.2.1)
Requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.7/dist-
packages (from gensim==3.8.1->-r requirements.txt (line 1)) (1.15.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-
packages (from spacy==2.2.3->-r requirements.txt (line 4)) (57.4.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (3.0.6)
Collecting srsly<1.1.0,>=0.1.0
    Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
      |                               | 184 kB 72.5 MB/s
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (1.0.7)
Collecting catalogue<1.1.0,>=0.0.7
    Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (2.23.0)
Collecting blis<0.5.0,>=0.4.0
    Downloading blis-0.4.1-cp37-cp37m-manylinux1_x86_64.whl (3.7 MB)
      |                               | 3.7 MB 57.4 MB/s
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (2.0.6)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (0.10.1)
Collecting thinc<7.4.0,>=7.3.0
    Downloading thinc-7.3.1-cp37-cp37m-manylinux1_x86_64.whl (2.2 MB)
      |                               | 2.2 MB 68.4 MB/s
Collecting plac<1.2.0,>=0.9.6
    Downloading plac-1.1.3-py2.py3-none-any.whl (20 kB)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/dist-
packages (from sklearn==0.0->-r requirements.txt (line 6)) (1.0.2)
Requirement already satisfied: protobuf>=3.8.0 in /usr/local/lib/python3.7/dist-
packages (from tensorboardX==2.0->-r requirements.txt (line 8)) (3.17.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-
packages (from transformers==3.1.0->-r requirements.txt (line 11)) (3.7.1)

```

```

Collecting sacremoses
  Downloading sacremoses-0.0.53.tar.gz (880 kB)
    |                               | 880 kB 57.1 MB/s
Collecting tokenizers==0.8.1.rc2
  Downloading tokenizers-0.8.1rc2-cp37-cp37m-manylinux1_x86_64.whl (3.0 MB)
    |                               | 3.0 MB 41.9 MB/s
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from transformers==3.1.0->-r requirements.txt (line 11)) (21.3)
Requirement already satisfied: importlib-metadata>=0.20 in /usr/local/lib/python3.7/dist-packages (from catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (4.12.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (4.1.1)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (1.24.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (2022.6.15)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging->transformers==3.1.0->-r requirements.txt (line 11)) (3.0.9)
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (7.1.2)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (1.1.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn->sklearn==0.0->-r requirements.txt (line 6)) (3.1.0)
Building wheels for collected packages: nltk, sacremoses
  Building wheel for nltk (setup.py) ... done
    Created wheel for nltk: filename=nltk-3.4.5-py3-none-any.whl size=1449923 sha256=1acebc2901ff32f3d22ded7faa01341787170c30795371f5af1a2faf43e1cd7e
    Stored in directory: /root/.cache/pip/wheels/48/8b/7f/473521e0c731c6566d631b281f323842bbda9bd819eb9a3ead
  Building wheel for sacremoses (setup.py) ... done

```

```
Created wheel for sacremoses: filename=sacremoses-0.0.53-py3-none-any.whl
size=895260
sha256=dabb19d8eeba09c2f2d7e4b2307bf41e14faac7e588b5b6ca3c4c2a4876bb960
Stored in directory: /root/.cache/pip/wheels/87/39/dd/a83eeef36d0bf98e7a4d1933
a4ad2d660295a40613079bafc9
Successfully built nltk sacremoses
Installing collected packages: numpy, tqdm, srsly, regex, plac, blis,
tokenizers, thinc, sentencepiece, sacremoses, catalogue, transformers, torch,
tensorboardX, spacy, nltk, jsonlines, gensim
Attempting uninstall: numpy
  Found existing installation: numpy 1.21.6
  Uninstalling numpy-1.21.6:
    Successfully uninstalled numpy-1.21.6
Attempting uninstall: tqdm
  Found existing installation: tqdm 4.64.0
  Uninstalling tqdm-4.64.0:
    Successfully uninstalled tqdm-4.64.0
Attempting uninstall: srsly
  Found existing installation: srsly 2.4.4
  Uninstalling srsly-2.4.4:
    Successfully uninstalled srsly-2.4.4
Attempting uninstall: regex
  Found existing installation: regex 2022.6.2
  Uninstalling regex-2022.6.2:
    Successfully uninstalled regex-2022.6.2
Attempting uninstall: blis
  Found existing installation: blis 0.7.8
  Uninstalling blis-0.7.8:
    Successfully uninstalled blis-0.7.8
Attempting uninstall: thinc
  Found existing installation: thinc 8.1.0
  Uninstalling thinc-8.1.0:
    Successfully uninstalled thinc-8.1.0
Attempting uninstall: catalogue
  Found existing installation: catalogue 2.0.8
  Uninstalling catalogue-2.0.8:
    Successfully uninstalled catalogue-2.0.8
Attempting uninstall: torch
  Found existing installation: torch 1.12.0+cu113
  Uninstalling torch-1.12.0+cu113:
    Successfully uninstalled torch-1.12.0+cu113
Attempting uninstall: spacy
  Found existing installation: spacy 3.4.1
  Uninstalling spacy-3.4.1:
    Successfully uninstalled spacy-3.4.1
Attempting uninstall: nltk
  Found existing installation: nltk 3.7
  Uninstalling nltk-3.7:
```

```

    Successfully uninstalled nltk-3.7
Attempting uninstall: gensim
    Found existing installation: gensim 3.6.0
    Uninstalling gensim-3.6.0:
        Successfully uninstalled gensim-3.6.0
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

xarray-einstats 0.2.2 requires numpy>=1.21, but you have numpy 1.19.2 which is
incompatible.

torchvision 0.13.0+cu113 requires torch==1.12.0, but you have torch 1.4.0 which
is incompatible.

torchtext 0.13.0 requires torch==1.12.0, but you have torch 1.4.0 which is
incompatible.

torchaudio 0.12.0+cu113 requires torch==1.12.0, but you have torch 1.4.0 which
is incompatible.

tensorflow 2.8.2+zzzcolab20220719082949 requires numpy>=1.20, but you have numpy
1.19.2 which is incompatible.

panel 0.12.1 requires tqdm>=4.48.0, but you have tqdm 4.41.0 which is
incompatible.

fastai 2.7.7 requires torch<1.13,>=1.7, but you have torch 1.4.0 which is
incompatible.

en-core-web-sm 3.4.0 requires spacy<3.5.0,>=3.4.0, but you have spacy 2.2.3
which is incompatible.

cmdstanpy 1.0.4 requires numpy>=1.21, but you have numpy 1.19.2 which is
incompatible.
Successfully installed blis-0.4.1 catalogue-1.0.0 gensim-3.8.1 jsonlines-1.2.0
nltk-3.4.5 numpy-1.19.2 plac-1.1.3 regex-2019.12.20 sacremoses-0.0.53
sentencepiece-0.1.85 spacy-2.2.3 srsly-1.0.5 tensorboardX-2.0 thinc-7.3.1
tokenizers-0.8.1rc2 torch-1.4.0 tqdm-4.41.0 transformers-3.1.0
/content/ditto

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

[ ]: True

```

## 0.2 Install fp16 optimization

```
[ ]: !git clone https://github.com/NVIDIA/apex
      %cd apex
      !pip install -v --no-cache-dir ./
      %cd ..
```

```
Cloning into 'apex'...
remote: Enumerating objects: 10085, done.
remote: Counting objects: 100% (201/201), done.
remote: Compressing objects: 100% (139/139), done.
remote: Total 10085 (delta 98), reused 130 (delta 58), pack-reused 9884
Receiving objects: 100% (10085/10085), 14.95 MiB | 7.45 MiB/s, done.
Resolving deltas: 100% (6903/6903), done.
/content/ditto/apex
Using pip 21.1.3 from /usr/local/lib/python3.7/dist-packages/pip (python 3.7)
Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m/UNKNOWN
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None
Non-user install because site-packages writable
Created temporary directory: /tmp/pip-ephem-wheel-cache-ytvwhjpe
Created temporary directory: /tmp/pip-req-tracker-t66eh5bu
Initialized build tracking at /tmp/pip-req-tracker-t66eh5bu
Created build tracker: /tmp/pip-req-tracker-t66eh5bu
Entered build tracker: /tmp/pip-req-tracker-t66eh5bu
```

```

Created temporary directory: /tmp/pip-install-7r0hl9n5
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Processing /content/ditto/apex
  Created temporary directory: /tmp/pip-req-build-c6qjo4d2
  DEPRECATION: A future pip version will change local packages to be built
in-place without first copying to a temporary directory. We recommend you use
--use-feature=in-tree-build to test your packages with this new behavior before
it becomes the default.

  pip 21.3 will remove support for this functionality. You can find discussion
regarding this at https://github.com/pypa/pip/issues/7555.
  Added file:///content/ditto/apex to build tracker '/tmp/pip-req-
tracker-t66eh5bu'
  Running setup.py (path:/tmp/pip-req-build-c6qjo4d2/setup.py) egg_info for
package from file:///content/ditto/apex
  Created temporary directory: /tmp/pip-pip-egg-info-29bw496l
  Running command python setup.py egg_info

  torch.__version__ = 1.4.0

  running egg_info
  creating /tmp/pip-pip-egg-info-29bw496l/apex.egg-info
  writing /tmp/pip-pip-egg-info-29bw496l/apex.egg-info/PKG-INFO
  writing dependency_links to /tmp/pip-pip-egg-info-29bw496l/apex.egg-
info/dependency_links.txt
  writing top-level names to /tmp/pip-pip-egg-info-29bw496l/apex.egg-
info/top_level.txt
  writing manifest file '/tmp/pip-pip-egg-info-29bw496l/apex.egg-
info/SOURCES.txt'
  adding license file 'LICENSE'
  writing manifest file '/tmp/pip-pip-egg-info-29bw496l/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-c6qjo4d2 has version 0.1, which satisfies
requirement apex==0.1 from file:///content/ditto/apex
  Removed apex==0.1 from file:///content/ditto/apex from build tracker
'/tmp/pip-req-tracker-t66eh5bu'
Created temporary directory: /tmp/pip-unpack-ivhga1k9
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-gbta22qs
  Building wheel for apex (setup.py) ... Destination directory:
/tmp/pip-wheel-gbta22qs
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = ''"/tmp/pip-req-build-c6qjo4d2/setup.py''';

```



```
__file__='''/tmp/pip-req-build-c6qjo4d2/setup.py''';f = getattr(tokenize,
''''open'''' , open)(__file__) if os.path.exists(__file__) else
io.StringIO('''from setuptools import setup; setup()''');code =
f.read().replace('''\r\n''', '''\n''');f.close();exec(compile(code,
__file__, '''exec'''))' bdist_wheel -d /tmp/pip-wheel-gbta22qs
```

```
torch.__version__ = 1.4.0
```

```
running bdist_wheel
running build
running build_py
creating build
creating build/lib
creating build/lib/apex
copying apex/__init__.py -> build/lib/apex
copying apex/_autocast_utils.py -> build/lib/apex
creating build/lib/apex/normalization
copying apex/normalization/__init__.py -> build/lib/apex/normalization
copying apex/normalization/fused_layer_norm.py -> build/lib/apex/normalization
creating build/lib/apex/RNN
copying apex/RNN/__init__.py -> build/lib/apex/RNN
copying apex/RNN/cells.py -> build/lib/apex/RNN
copying apex/RNN/RNNBackend.py -> build/lib/apex/RNN
copying apex/RNN/models.py -> build/lib/apex/RNN
creating build/lib/apex/contrib
copying apex/contrib/__init__.py -> build/lib/apex/contrib
creating build/lib/apex/multi_tensor_apply
copying apex/multi_tensor_apply/__init__.py ->
build/lib/apex/multi_tensor_apply
copying apex/multi_tensor_apply/multi_tensor_apply.py ->
build/lib/apex/multi_tensor_apply
creating build/lib/apex/transformer
copying apex/transformer/__init__.py -> build/lib/apex/transformer
copying apex/transformer/enums.py -> build/lib/apex/transformer
copying apex/transformer/microbatches.py -> build/lib/apex/transformer
copying apex/transformer/log_util.py -> build/lib/apex/transformer
copying apex/transformer/utils.py -> build/lib/apex/transformer
copying apex/transformer/parallel_state.py -> build/lib/apex/transformer
creating build/lib/apex/amp
copying apex/amp/__init__.py -> build/lib/apex/amp
copying apex/amp/opt.py -> build/lib/apex/amp
copying apex/amp/rnn_compat.py -> build/lib/apex/amp
copying apex/amp/_process_optimizer.py -> build/lib/apex/amp
copying apex/amp/compat.py -> build/lib/apex/amp
copying apex/amp/__version__.py -> build/lib/apex/amp
copying apex/amp/utils.py -> build/lib/apex/amp
```

```

copying apex/amp/scaler.py -> build/lib/apex/amp
copying apex/amp/wrap.py -> build/lib/apex/amp
copying apex/amp/frontend.py -> build/lib/apex/amp
copying apex/amp/_initialize.py -> build/lib/apex/amp
copying apex/amp/handle.py -> build/lib/apex/amp
copying apex/amp/_amp_state.py -> build/lib/apex/amp
copying apex/amp/amp.py -> build/lib/apex/amp
creating build/lib/apex/optimizers
copying apex/optimizers/__init__.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_sgd.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_mixed_precision_lamb.py ->
build/lib/apex/optimizers
copying apex/optimizers/fused_novograd.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_adagrad.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_adam.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_lamb.py -> build/lib/apex/optimizers
creating build/lib/apex/fp16_utils
copying apex/fp16_utils/__init__.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/loss_scaler.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/fp16util.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/fp16_optimizer.py -> build/lib/apex/fp16_utils
creating build/lib/apex/fused_dense
copying apex/fused_dense/__init__.py -> build/lib/apex/fused_dense
copying apex/fused_dense/fused_dense.py -> build/lib/apex/fused_dense
creating build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm.py -> build/lib/apex/parallel
copying apex/parallel/__init__.py -> build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/lib/apex/parallel
copying apex/parallel/sync_batchnorm_kernel.py -> build/lib/apex/parallel
copying apex/parallel/multiproc.py -> build/lib/apex/parallel
copying apex/parallel/distributed.py -> build/lib/apex/parallel
copying apex/parallel/LARC.py -> build/lib/apex/parallel
copying apex/parallel/sync_batchnorm.py -> build/lib/apex/parallel
creating build/lib/apex/mlp
copying apex/mlp/__init__.py -> build/lib/apex/mlp
copying apex/mlp/mlp.py -> build/lib/apex/mlp
creating build/lib/apex/contrib/focal_loss
copying apex/contrib/focal_loss/__init__.py ->
build/lib/apex/contrib/focal_loss
copying apex/contrib/focal_loss/focal_loss.py ->
build/lib/apex/contrib/focal_loss
creating build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/__init__.py ->
build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/peer_halo_exchanger_1d.py ->
build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/peer_memory.py ->

```

```

build/lib/apex/contrib/peer_memory
  copying apex/contrib/peer_memory/peer_halo_exchange_module_tests.py ->
build/lib/apex/contrib/peer_memory
  creating build/lib/apex/contrib/groupbn
  copying apex/contrib/groupbn/__init__.py -> build/lib/apex/contrib/groupbn
  copying apex/contrib/groupbn/batch_norm.py -> build/lib/apex/contrib/groupbn
  creating build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/__init__.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/distributed_fused_adam.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_sgd.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_adam.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_lamb.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/distributed_fused_lamb.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fp16_optimizer.py ->
build/lib/apex/contrib/optimizers
  creating build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/__init__.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/self_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py
-> build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/fast_self_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  copying
apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_add_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/self_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  creating build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/__init__.py ->
build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/bottleneck_module_test.py ->
build/lib/apex/contrib/bottleneck

```

```

    copying apex/contrib/bottleneck/test.py -> build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/bottleneck.py ->
build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/halo_exchangers.py ->
build/lib/apex/contrib/bottleneck
    creating build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/__init__.py -> build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/permutation_lib.py ->
build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/asp.py -> build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/sparse_masklib.py ->
build/lib/apex/contrib/sparsity
    creating build/lib/apex/contrib/layer_norm
    copying apex/contrib/layer_norm/__init__.py ->
build/lib/apex/contrib/layer_norm
    copying apex/contrib/layer_norm/layer_norm.py ->
build/lib/apex/contrib/layer_norm
    creating build/lib/apex/contrib/transducer
    copying apex/contrib/transducer/__init__.py ->
build/lib/apex/contrib/transducer
    copying apex/contrib/transducer/transducer.py ->
build/lib/apex/contrib/transducer
    creating build/lib/apex/contrib/index_mul_2d
    copying apex/contrib/index_mul_2d/__init__.py ->
build/lib/apex/contrib/index_mul_2d
    copying apex/contrib/index_mul_2d/index_mul_2d.py ->
build/lib/apex/contrib/index_mul_2d
    creating build/lib/apex/contrib/conv_bias_relu
    copying apex/contrib/conv_bias_relu/__init__.py ->
build/lib/apex/contrib/conv_bias_relu
    copying apex/contrib/conv_bias_relu/conv_bias_relu.py ->
build/lib/apex/contrib/conv_bias_relu
    creating build/lib/apex/contrib/xentropy
    copying apex/contrib/xentropy/__init__.py -> build/lib/apex/contrib/xentropy
    copying apex/contrib/xentropy/softmax_xentropy.py ->
build/lib/apex/contrib/xentropy
    creating build/lib/apex/contrib/clip_grad
    copying apex/contrib/clip_grad/__init__.py -> build/lib/apex/contrib/clip_grad
    copying apex/contrib/clip_grad/clip_grad.py ->
build/lib/apex/contrib/clip_grad
    creating build/lib/apex/contrib/fmha
    copying apex/contrib/fmha/__init__.py -> build/lib/apex/contrib/fmha
    copying apex/contrib/fmha/fmha.py -> build/lib/apex/contrib/fmha
    creating build/lib/apex/contrib/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/__init__.py ->
build/lib/apex/contrib/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/call_permutation_sear
ch_kernels.py -> build/lib/apex/contrib/sparsity/permutation_search_kernels

```

```

copying
apex/contrib/sparsity/permutation_search_kernels/permutation_utilities.py ->
build/lib/apex/contrib/sparsity/permutation_search_kernels
copying apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py
-> build/lib/apex/contrib/sparsity/permutation_search_kernels
creating build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/__init__.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/mappings.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/layers.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/memory.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/utils.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/cross_entropy.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/random.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/data.py ->
build/lib/apex/transformer/tensor_parallel
creating build/lib/apex/transformer/testing
copying apex/transformer/testing/__init__.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/distributed_test_base.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/standalone_bert.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/arguments.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/standalone_transformer_lm.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/commons.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/global_vars.py ->
build/lib/apex/transformer/testing
copying apex/transformer/testing/standalone_gpt.py ->
build/lib/apex/transformer/testing
creating build/lib/apex/transformer/functional
copying apex/transformer/functional/__init__.py ->
build/lib/apex/transformer/functional
copying apex/transformer/functional/fused_softmax.py ->
build/lib/apex/transformer/functional
creating build/lib/apex/transformer/amp
copying apex/transformer/amp/__init__.py -> build/lib/apex/transformer/amp
copying apex/transformer/amp/grad_scaler.py -> build/lib/apex/transformer/amp
creating build/lib/apex/transformer/_data

```

```

    copying apex/transformer/_data/__init__.py -> build/lib/apex/transformer/_data
    copying apex/transformer/_data/_batch_sampler.py ->
build/lib/apex/transformer/_data
    creating build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/__init__.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/_timers.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/utils.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/p2p_communication.py ->
build/lib/apex/transformer/pipeline_parallel
    creating build/lib/apex/transformer/layers
    copying apex/transformer/layers/__init__.py ->
build/lib/apex/transformer/layers
    copying apex/transformer/layers/layer_norm.py ->
build/lib/apex/transformer/layers
    creating build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_with_interleaving.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/__init__.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipeline_parallel.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/common.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_without_interleaving.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    creating build/lib/apex/amp/lists
    copying apex/amp/lists/__init__.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/tensor_overrides.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/functional_overrides.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/torch_overrides.py -> build/lib/apex/amp/lists
    installing to build/bdist.linux-x86_64/wheel
    running install
    running install_lib
    creating build/bdist.linux-x86_64
    creating build/bdist.linux-x86_64/wheel
    creating build/bdist.linux-x86_64/wheel/apex
    copying build/lib/apex/__init__.py -> build/bdist.linux-x86_64/wheel/apex
    copying build/lib/apex/_autocast_utils.py ->
build/bdist.linux-x86_64/wheel/apex
    creating build/bdist.linux-x86_64/wheel/apex/normalization
    copying build/lib/apex/normalization/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/normalization
    copying build/lib/apex/normalization/fused_layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/normalization
    creating build/bdist.linux-x86_64/wheel/apex/RNN

```

```

    copying build/lib/apex/RNN/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/cells.py -> build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/RNNBackend.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/models.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    creating build/bdist.linux-x86_64/wheel/apex/contrib
    creating build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/focal_loss/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/focal_loss/focal_loss.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib
    creating build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_halo_exchanger_1d.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_memory.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_halo_exchange_module_tests.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    creating build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    copying build/lib/apex/contrib/groupbn/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    copying build/lib/apex/contrib/groupbn/batch_norm.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    creating build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/distributed_fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_sgd.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/distributed_fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fp16_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    creating build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/self_multihead_attn.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying
build/lib/apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/fast_self_multihead_attn_func.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying
build/lib/apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_
add_func.py -> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/self_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    creating build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/bottleneck_module_test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/bottleneck.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/halo_exchangers.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    creating build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/permutation_lib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/asp.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/sparse_masklib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    creating
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/__init__.py
->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/call_permut
ation_search_kernels.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/permutation

```



```

_utilities.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying
build/lib/apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py
->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    creating build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    copying build/lib/apex/contrib/layer_norm/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    copying build/lib/apex/contrib/layer_norm/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    creating build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    copying build/lib/apex/contrib/transducer/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    copying build/lib/apex/contrib/transducer/transducer.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    creating build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    copying build/lib/apex/contrib/index_mul_2d/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    copying build/lib/apex/contrib/index_mul_2d/index_mul_2d.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    creating build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    copying build/lib/apex/contrib/conv_bias_relu/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    copying build/lib/apex/contrib/conv_bias_relu/conv_bias_relu.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    creating build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    copying build/lib/apex/contrib/xentropy/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    copying build/lib/apex/contrib/xentropy/softmax_xentropy.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    creating build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    copying build/lib/apex/contrib/clip_grad/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    copying build/lib/apex/contrib/clip_grad/clip_grad.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    creating build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    copying build/lib/apex/contrib/fmha/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    copying build/lib/apex/contrib/fmha/fmha.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    creating build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    copying build/lib/apex/multi_tensor_apply/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    copying build/lib/apex/multi_tensor_apply/multi_tensor_apply.py ->
build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    creating build/bdist.linux-x86_64/wheel/apex/transformer
    copying build/lib/apex/transformer/__init__.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/mappings.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/layers.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/memory.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/cross_entropy.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/random.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/data.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/enums.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  copying build/lib/apex/transformer/microbatches.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/distributed_test_base.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_bert.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/arguments.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_transformer_lm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/commons.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/global_vars.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_gpt.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  creating build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/fused_softmax.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/log_util.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/amp/__init__.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/amp/grad_scaler.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/_data
  copying build/lib/apex/transformer/_data/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/_data
  copying build/lib/apex/transformer/_data/_batch_sampler.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/_data
  creating build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/_timers.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  creating
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_with_interleaving.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying
build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipeline_parallel.py
-> build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/common.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_without_interleaving.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/p2p_communication.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  creating build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/parallel_state.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/opt.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/rnn_compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_process_optimizer.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  creating build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/tensor_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/functional_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/torch_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/__version__.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/utils.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/scaler.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/wrap.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/frontend.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_initialize.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/handle.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_amp_state.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/amp.py -> build/bdist.linux-x86_64/wheel/apex/amp
  creating build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_sgd.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_mixed_precision_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_novograd.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_adagrad.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  creating build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/loss_scaler.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16util.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  creating build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/fused_dense.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  creating build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/multiproc.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/distributed.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/LARC.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  creating build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/mlp.py -> build/bdist.linux-x86_64/wheel/apex/mlp
  running install_egg_info
  running egg_info
  creating apex.egg-info
  writing apex.egg-info/PKG-INFO
  writing dependency_links to apex.egg-info/dependency_links.txt
  writing top-level names to apex.egg-info/top_level.txt
  writing manifest file 'apex.egg-info/SOURCES.txt'
  adding license file 'LICENSE'
  writing manifest file 'apex.egg-info/SOURCES.txt'
  Copying apex.egg-info to build/bdist.linux-x86_64/wheel/apex-0.1-py3.7.egg-
info
  running install_scripts
  adding license file "LICENSE" (matched pattern "LICEN[CS]E*")
  creating build/bdist.linux-x86_64/wheel/apex-0.1.dist-info/WHEEL
  creating '/tmp/pip-wheel-gbta22qs/apex-0.1-py3-none-any.whl' and adding
'build/bdist.linux-x86_64/wheel' to it
  adding 'apex/__init__.py'
  adding 'apex/_autocast_utils.py'
  adding 'apex/RNN/RNNBackend.py'

```

```

adding 'apex/RNN/__init__.py'
adding 'apex/RNN/cells.py'
adding 'apex/RNN/models.py'
adding 'apex/amp/__init__.py'
adding 'apex/amp/__version__.py'
adding 'apex/amp/_amp_state.py'
adding 'apex/amp/_initialize.py'
adding 'apex/amp/_process_optimizer.py'
adding 'apex/amp/amp.py'
adding 'apex/amp/compat.py'
adding 'apex/amp/frontend.py'
adding 'apex/amp/handle.py'
adding 'apex/amp/opt.py'
adding 'apex/amp/rnn_compat.py'
adding 'apex/amp/scaler.py'
adding 'apex/amp/utils.py'
adding 'apex/amp/wrap.py'
adding 'apex/amp/lists/__init__.py'
adding 'apex/amp/lists/functional_overrides.py'
adding 'apex/amp/lists/tensor_overrides.py'
adding 'apex/amp/lists/torch_overrides.py'
adding 'apex/contrib/__init__.py'
adding 'apex/contrib/bottleneck/__init__.py'
adding 'apex/contrib/bottleneck/bottleneck.py'
adding 'apex/contrib/bottleneck/bottleneck_module_test.py'
adding 'apex/contrib/bottleneck/halo_exchangers.py'
adding 'apex/contrib/bottleneck/test.py'
adding 'apex/contrib/clip_grad/__init__.py'
adding 'apex/contrib/clip_grad/clip_grad.py'
adding 'apex/contrib/conv_bias_relu/__init__.py'
adding 'apex/contrib/conv_bias_relu/conv_bias_relu.py'
adding 'apex/contrib/fmha/__init__.py'
adding 'apex/contrib/fmha/fmha.py'
adding 'apex/contrib/focal_loss/__init__.py'
adding 'apex/contrib/focal_loss/focal_loss.py'
adding 'apex/contrib/groupbn/__init__.py'
adding 'apex/contrib/groupbn/batch_norm.py'
adding 'apex/contrib/index_mul_2d/__init__.py'
adding 'apex/contrib/index_mul_2d/index_mul_2d.py'
adding 'apex/contrib/layer_norm/__init__.py'
adding 'apex/contrib/layer_norm/layer_norm.py'
adding 'apex/contrib/multihead_attn/__init__.py'
adding 'apex/contrib/multihead_attn/encdec_multihead_attn.py'
adding 'apex/contrib/multihead_attn/encdec_multihead_attn_func.py'
adding 'apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py'
adding
'apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_add_func.py'
    adding 'apex/contrib/multihead_attn/fast_self_multihead_attn_func.py'

```

```

adding 'apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py'
adding 'apex/contrib/multihead_attn/mask_softmax_dropout_func.py'
adding 'apex/contrib/multihead_attn/self_multihead_attn.py'
adding 'apex/contrib/multihead_attn/self_multihead_attn_func.py'
adding 'apex/contrib/optimizers/__init__.py'
adding 'apex/contrib/optimizers/distributed_fused_adam.py'
adding 'apex/contrib/optimizers/distributed_fused_lamb.py'
adding 'apex/contrib/optimizers/fp16_optimizer.py'
adding 'apex/contrib/optimizers/fused_adam.py'
adding 'apex/contrib/optimizers/fused_lamb.py'
adding 'apex/contrib/optimizers/fused_sgd.py'
adding 'apex/contrib/peer_memory/__init__.py'
adding 'apex/contrib/peer_memory/peer_halo_exchange_module_tests.py'
adding 'apex/contrib/peer_memory/peer_halo_exchanger_1d.py'
adding 'apex/contrib/peer_memory/peer_memory.py'
adding 'apex/contrib/sparsity/__init__.py'
adding 'apex/contrib/sparsity/asp.py'
adding 'apex/contrib/sparsity/permutation_lib.py'
adding 'apex/contrib/sparsity/sparse_masklib.py'
adding 'apex/contrib/sparsity/permutation_search_kernels/__init__.py'
adding 'apex/contrib/sparsity/permutation_search_kernels/call_permutation_sear
ch_kernels.py'
    adding 'apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py'
    adding
'apex/contrib/sparsity/permutation_search_kernels/permutation_utilities.py'
    adding 'apex/contrib/transducer/__init__.py'
    adding 'apex/contrib/transducer/transducer.py'
    adding 'apex/contrib/xentropy/__init__.py'
    adding 'apex/contrib/xentropy/softmax_xentropy.py'
    adding 'apex/fp16_utils/__init__.py'
    adding 'apex/fp16_utils/fp16_optimizer.py'
    adding 'apex/fp16_utils/fp16util.py'
    adding 'apex/fp16_utils/loss_scaler.py'
    adding 'apex/fused_dense/__init__.py'
    adding 'apex/fused_dense/fused_dense.py'
    adding 'apex/mlp/__init__.py'
    adding 'apex/mlp/mlp.py'
    adding 'apex/multi_tensor_apply/__init__.py'
    adding 'apex/multi_tensor_apply/multi_tensor_apply.py'
    adding 'apex/normalization/__init__.py'
    adding 'apex/normalization/fused_layer_norm.py'
    adding 'apex/optimizers/__init__.py'
    adding 'apex/optimizers/fused_adagrad.py'
    adding 'apex/optimizers/fused_adam.py'
    adding 'apex/optimizers/fused_lamb.py'
    adding 'apex/optimizers/fused_mixed_precision_lamb.py'
    adding 'apex/optimizers/fused_novograd.py'
    adding 'apex/optimizers/fused_sgd.py'

```

```

adding 'apex/parallel/LARC.py'
adding 'apex/parallel/__init__.py'
adding 'apex/parallel/distributed.py'
adding 'apex/parallel/multiproc.py'
adding 'apex/parallel/optimized_sync_batchnorm.py'
adding 'apex/parallel/optimized_sync_batchnorm_kernel.py'
adding 'apex/parallel/sync_batchnorm.py'
adding 'apex/parallel/sync_batchnorm_kernel.py'
adding 'apex/transformer/__init__.py'
adding 'apex/transformer/enums.py'
adding 'apex/transformer/log_util.py'
adding 'apex/transformer/microbatches.py'
adding 'apex/transformer/parallel_state.py'
adding 'apex/transformer/utils.py'
adding 'apex/transformer/_data/__init__.py'
adding 'apex/transformer/_data/_batch_sampler.py'
adding 'apex/transformer/amp/__init__.py'
adding 'apex/transformer/amp/grad_scaler.py'
adding 'apex/transformer/functional/__init__.py'
adding 'apex/transformer/functional/fused_softmax.py'
adding 'apex/transformer/layers/__init__.py'
adding 'apex/transformer/layers/layer_norm.py'
adding 'apex/transformer/pipeline_parallel/__init__.py'
adding 'apex/transformer/pipeline_parallel/_timers.py'
adding 'apex/transformer/pipeline_parallel/p2p_communication.py'
adding 'apex/transformer/pipeline_parallel/utils.py'
adding 'apex/transformer/pipeline_parallel/schedules/__init__.py'
adding 'apex/transformer/pipeline_parallel/schedules/common.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipelining.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_with_i
nterleaving.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_withou
t_interleaving.py'
adding 'apex/transformer/tensor_parallel/__init__.py'
adding 'apex/transformer/tensor_parallel/cross_entropy.py'
adding 'apex/transformer/tensor_parallel/data.py'
adding 'apex/transformer/tensor_parallel/layers.py'
adding 'apex/transformer/tensor_parallel/mappings.py'
adding 'apex/transformer/tensor_parallel/memory.py'
adding 'apex/transformer/tensor_parallel/random.py'
adding 'apex/transformer/tensor_parallel/utils.py'
adding 'apex/transformer/testing/__init__.py'
adding 'apex/transformer/testing/arguments.py'
adding 'apex/transformer/testing/commons.py'
adding 'apex/transformer/testing/distributed_test_base.py'
adding 'apex/transformer/testing/global_vars.py'
adding 'apex/transformer/testing/standalone_bert.py'
adding 'apex/transformer/testing/standalone_gpt.py'

```



```

adding 'apex/transformer/testing/standalone_transformer_lm.py'
adding 'apex-0.1.dist-info/LICENSE'
adding 'apex-0.1.dist-info/METADATA'
adding 'apex-0.1.dist-info/WHEEL'
adding 'apex-0.1.dist-info/top_level.txt'
adding 'apex-0.1.dist-info/RECORD'
removing build/bdist.linux-x86_64/wheel
done
Created wheel for apex: filename=apex-0.1-py3-none-any.whl size=298114
sha256=4b3b56aad897b46a4167abbd6f553a3a12b95baef7c88543b0a19122a5506aa1
Stored in directory: /tmp/pip-ephem-wheel-cache-
ytvwhjpe/wheels/6b/22/a2/1665526ee3c3061243260fd522525eca31398e04bfa5ad7e1f
Successfully built apex
Installing collected packages: apex
Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/apex
sysconfig: /usr/include/python3.7m/apex
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None

Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages

```

```

Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m/UNKNOWN
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None
Successfully installed apex-0.1
Removed build tracker: '/tmp/pip-req-tracker-t66eh5bu'
/content/ditto

```

```

[ ]: # some issue with colab
!pip install --upgrade "urllib3==1.25.4" awscli

```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting urllib3==1.25.4
  Downloading urllib3-1.25.4-py2.py3-none-any.whl (125 kB)
    |                               | 125 kB 24.0 MB/s
Collecting awscli
  Downloading awscli-1.25.46-py3-none-any.whl (3.9 MB)
    |                               | 3.9 MB 57.2 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
    |                               | 79 kB 10.3 MB/s
Collecting docutils<0.17,>=0.10
  Downloading docutils-0.16-py2.py3-none-any.whl (548 kB)
    |                               | 548 kB 76.1 MB/s
Requirement already satisfied: PyYAML<5.5,>=3.10 in
/usr/local/lib/python3.7/dist-packages (from awscli) (3.13)
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting botocore==1.27.46
  Downloading botocore-1.27.46-py3-none-any.whl (9.0 MB)
    |                               | 9.0 MB 56.3 MB/s
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in

```

```

/usr/local/lib/python3.7/dist-packages (from botocore==1.27.46->awscli) (2.8.2)
Collecting jmespath<2.0.0,>=0.7.1
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil<3.0.0,>=2.1->botocore==1.27.46->awscli) (1.15.0)
Requirement already satisfied: pyasn1>=0.1.3 in /usr/local/lib/python3.7/dist-
packages (from rsa<4.8,>=3.1.2->awscli) (0.4.8)
Installing collected packages: urllib3, jmespath, botocore, s3transfer, rsa,
docutils, colorama, awscli
  Attempting uninstall: urllib3
    Found existing installation: urllib3 1.24.3
    Uninstalling urllib3-1.24.3:
      Successfully uninstalled urllib3-1.24.3
  Attempting uninstall: rsa
    Found existing installation: rsa 4.9
    Uninstalling rsa-4.9:
      Successfully uninstalled rsa-4.9
  Attempting uninstall: docutils
    Found existing installation: docutils 0.17.1
    Uninstalling docutils-0.17.1:
      Successfully uninstalled docutils-0.17.1
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

panel 0.12.1 requires tqdm>=4.48.0, but you have tqdm 4.41.0 which is
incompatible.

fastai 2.7.7 requires torch<1.13,>=1.7, but you have torch 1.4.0 which is
incompatible.

Successfully installed awscli-1.25.46 botocore-1.27.46 colorama-0.4.4
docutils-0.16 jmespath-1.0.1 rsa-4.7.2 s3transfer-0.6.0 urllib3-1.25.4

```

```
[ ]: !pip install torch==1.7.1
```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting torch==1.7.1
  Downloading torch-1.7.1-cp37-cp37m-manylinux1_x86_64.whl (776.8 MB)
    |                                     | 776.8 MB 18 kB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-
packages (from torch==1.7.1) (1.19.2)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
Installing collected packages: torch
  Attempting uninstall: torch
    Found existing installation: torch 1.4.0

```

```

Uninstalling torch-1.4.0:
  Successfully uninstalled torch-1.4.0
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

torchvision 0.13.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.

torchtext 0.13.0 requires torch==1.12.0, but you have torch 1.7.1 which is
incompatible.

torchaudio 0.12.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.
Successfully installed torch-1.7.1

```

### 0.3 importa config.json

```
[ ]: !git clone https://github.com/pauloh48/IC.git
```

```

Cloning into 'IC'...
remote: Enumerating objects: 982, done.
remote: Counting objects: 100% (52/52), done.
remote: Compressing objects: 100% (48/48), done.
remote: Total 982 (delta 16), reused 7 (delta 0), pack-reused 930
Receiving objects: 100% (982/982), 251.89 MiB | 15.91 MiB/s, done.
Resolving deltas: 100% (600/600), done.
Checking out files: 100% (1273/1273), done.

```

### 0.4 remove e move

```
[ ]: !ls
```

```

apex          ditto.jpg    LICENSE     requirements.txt  run_all_wdc.py
blocking      ditto_light matcher.py  results_ditto    Snippext_public
configs.json  IC          output      run_all_er_magellan.py train_ditto.py
data          input       README.md   run_all_vary_size.py

```

```
[ ]: !rm configs.json
!mv -n /content/ditto/IC/datasEsErros/GLow_30_1_2/configs.json /content/ditto/
```

# 1 DA

## 1.1 DA STRUCTURED

### 1.1.1 Beer - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Beer \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --da attr_del \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

```
Downloading: 100% 442/442 [00:00<00:00, 375kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 349kB/s]
Downloading: 100% 268M/268M [00:06<00:00, 39.2MB/s]
step: 0, loss: 0.8852607607841492
epoch 1: dev_f1=0.30434782608695654, f1=0.27184466019417475,
best_f1=0.27184466019417475
step: 0, loss: 0.3630201816558838
epoch 2: dev_f1=0.288659793814433, f1=0.27184466019417475,
best_f1=0.27184466019417475
step: 0, loss: 0.33076441287994385
epoch 3: dev_f1=0.38095238095238093, f1=0.34285714285714286,
best_f1=0.34285714285714286
```

```

step: 0, loss: 0.37268945574760437
epoch 4: dev_f1=0.39285714285714285, f1=0.30000000000000004,
best_f1=0.30000000000000004
step: 0, loss: 0.28805944323539734
epoch 5: dev_f1=0.39285714285714285, f1=0.31168831168831174,
best_f1=0.30000000000000004
step: 0, loss: 0.3194631040096283
epoch 6: dev_f1=0.42857142857142855, f1=0.32000000000000006,
best_f1=0.32000000000000006
step: 0, loss: 0.2588895857334137
epoch 7: dev_f1=0.48148148148148157, f1=0.3561643835616438,
best_f1=0.3561643835616438
step: 0, loss: 0.3555213510990143
epoch 8: dev_f1=0.4827586206896552, f1=0.4210526315789474,
best_f1=0.4210526315789474
step: 0, loss: 0.23101870715618134
epoch 9: dev_f1=0.5, f1=0.38095238095238093, best_f1=0.38095238095238093
step: 0, loss: 0.2258005142211914
epoch 10: dev_f1=0.5714285714285714, f1=0.3333333333333333,
best_f1=0.3333333333333333
step: 0, loss: 0.19891755282878876
epoch 11: dev_f1=0.6086956521739131, f1=0.45161290322580644,
best_f1=0.45161290322580644
step: 0, loss: 0.19273671507835388
epoch 12: dev_f1=0.5833333333333334, f1=0.4242424242424242,
best_f1=0.45161290322580644
step: 0, loss: 0.09453975409269333
epoch 13: dev_f1=0.6086956521739131, f1=0.4242424242424242,
best_f1=0.45161290322580644
step: 0, loss: 0.1572776436805725
epoch 14: dev_f1=0.6086956521739131, f1=0.4242424242424242,
best_f1=0.45161290322580644
step: 0, loss: 0.23560799658298492
epoch 15: dev_f1=0.6086956521739131, f1=0.4242424242424242,
best_f1=0.45161290322580644
mv: cannot stat '*_dev.pt': No such file or directory
Traceback (most recent call last):
  File "matcher.py", line 315, in <module>
    hp.lm, hp.use_gpu, hp.fp16)
  File "matcher.py", line 274, in load_model
    raise ModelNotFoundError(checkpoint)
ditto_light.exceptions.ModelNotFoundError: Model
checkpoints/Structured/Beer/model.pt was not found

```

### 1.1.2 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap \
  --da span_shuffle \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled           : True
```

```

opt_level          : 02
cast_model_type    : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights     : True
loss_scale         : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8156179189682007
step: 10, loss: 0.4497043490409851
step: 20, loss: 0.5751700401306152
step: 30, loss: 0.3574075996875763
step: 40, loss: 0.15982507169246674
step: 50, loss: 0.04221222549676895
step: 60, loss: 0.09964247792959213
step: 70, loss: 0.06619756668806076
step: 80, loss: 0.11464802175760269
step: 90, loss: 0.04722357913851738
step: 100, loss: 0.016565585508942604
step: 110, loss: 0.013799849897623062
step: 120, loss: 0.016065282747149467
step: 130, loss: 0.006625621113926172
step: 140, loss: 0.17262379825115204
step: 150, loss: 0.04167254641652107
step: 160, loss: 0.12017360329627991
step: 170, loss: 0.0023367919493466616
step: 180, loss: 0.011882726103067398
step: 190, loss: 0.03748970851302147
step: 200, loss: 0.007864214479923248
step: 210, loss: 0.007546708453446627
step: 220, loss: 0.0036777725908905268
step: 230, loss: 0.08136703819036484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9797752808988766, f1=0.9737742303306728,
best_f1=0.9737742303306728
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01146064605563879
step: 10, loss: 0.0009774381760507822
step: 20, loss: 0.0056159766390919685
step: 30, loss: 0.0017006989801302552
step: 40, loss: 0.007415245287120342
step: 50, loss: 0.036308206617832184
step: 60, loss: 0.0037612018641084433
step: 70, loss: 0.14277879893779755
step: 80, loss: 0.02020445466041565
step: 90, loss: 0.028143921867012978
step: 100, loss: 0.02395518310368061
step: 110, loss: 0.0071875168941915035
step: 120, loss: 0.004815331194549799
step: 130, loss: 0.003151179989799857
step: 140, loss: 0.13177111744880676
step: 150, loss: 0.012539413757622242
step: 160, loss: 0.0035469287540763617
step: 170, loss: 0.010728604160249233
step: 180, loss: 0.0017269126838073134
step: 190, loss: 0.0654686987400055
step: 200, loss: 0.0013528082054108381
step: 210, loss: 0.05724063515663147
step: 220, loss: 0.0004707033804152161
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 230, loss: 0.4434232711791992
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.967741935483871, f1=0.9596309111880046,
best_f1=0.9737742303306728
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21019016206264496
step: 10, loss: 0.053083740174770355
step: 20, loss: 0.03517074137926102
step: 30, loss: 0.020836643874645233

```

```

step: 40, loss: 0.010928090661764145
step: 50, loss: 0.006285019684582949
step: 60, loss: 0.12670797109603882
step: 70, loss: 0.000986986793577671
step: 80, loss: 0.0036290851421654224
step: 90, loss: 0.0012910838704556227
step: 100, loss: 0.0015361859695985913
step: 110, loss: 0.0024329591542482376
step: 120, loss: 0.001363550778478384
step: 130, loss: 0.0005889266612939537
step: 140, loss: 0.0005495423101820052
step: 150, loss: 0.001258500968106091
step: 160, loss: 0.003568737767636776
step: 170, loss: 0.004766850266605616
step: 180, loss: 0.0009717209031805396
step: 190, loss: 0.0014002136886119843
step: 200, loss: 0.013113162480294704
step: 210, loss: 0.061545077711343765
step: 220, loss: 0.0025176056660711765
step: 230, loss: 0.1030845120549202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9853768278965129, f1=0.984090909090909,
best_f1=0.984090909090909
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002062159590423107
step: 10, loss: 0.0037567082326859236
step: 20, loss: 0.00047231162898242474
step: 30, loss: 0.00038945217966102064
step: 40, loss: 0.008779173716902733
step: 50, loss: 0.0029505465645343065
step: 60, loss: 0.0006228717393241823
step: 70, loss: 0.002715387847274542
step: 80, loss: 0.10193409025669098
step: 90, loss: 0.0074014561250805855
step: 100, loss: 0.0008291050908155739
step: 110, loss: 0.010145195759832859
step: 120, loss: 0.061658505350351334
step: 130, loss: 0.0017998130060732365
step: 140, loss: 0.04836026951670647
step: 150, loss: 0.003286317689344287

```

```

step: 160, loss: 0.0005583903403021395
step: 170, loss: 0.006857973523437977
step: 180, loss: 0.04429345950484276
step: 190, loss: 0.005976373795419931
step: 200, loss: 0.0005666571669280529
step: 210, loss: 0.0004316802369430661
step: 220, loss: 0.0023029393050819635
step: 230, loss: 0.00039599306182935834
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9887640449438202, f1=0.9788182831661093,
best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007878721808083355
step: 10, loss: 0.0009584420477040112
step: 20, loss: 0.0010694474913179874
step: 30, loss: 0.0009194780141115189
step: 40, loss: 0.00038523622788488865
step: 50, loss: 0.014479557052254677
step: 60, loss: 0.0002820114605128765
step: 70, loss: 0.00015768477169331163
step: 80, loss: 0.0002781423681881279
step: 90, loss: 0.000209636491490528
step: 100, loss: 0.0009226181427948177
step: 110, loss: 0.0012352931080386043
step: 120, loss: 0.0370013602077961
step: 130, loss: 0.00455764913931489
step: 140, loss: 0.00017258689331356436
step: 150, loss: 0.0001864885853137821
step: 160, loss: 0.0022224620915949345
step: 170, loss: 0.0030928717460483313
step: 180, loss: 0.00032025473774410784
step: 190, loss: 0.0004590678436215967
step: 200, loss: 0.0026296149007976055
step: 210, loss: 0.00040292684570886195
step: 220, loss: 0.000527137890458107
step: 230, loss: 0.00040309413452632725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9887640449438202, f1=0.9755011135857461,

```

```

best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004577977815642953
step: 10, loss: 0.005353208165615797
step: 20, loss: 0.00033311540028080344
step: 30, loss: 0.000600429018959403
step: 40, loss: 0.0012151992414146662
step: 50, loss: 0.02791161835193634
step: 60, loss: 0.018475357443094254
step: 70, loss: 0.0002986832696478814
step: 80, loss: 0.0018766939174383879
step: 90, loss: 0.005823792889714241
step: 100, loss: 0.0015830849297344685
step: 110, loss: 0.039813753217458725
step: 120, loss: 0.0002907812304329127
step: 130, loss: 0.0016153239412233233
step: 140, loss: 0.021692877635359764
step: 150, loss: 0.0005139551940374076
step: 160, loss: 0.00026167716714553535
step: 170, loss: 0.0005483949207700789
step: 180, loss: 0.0003498656442388892
step: 190, loss: 0.0004135706985834986
step: 200, loss: 0.00041203893488273025
step: 210, loss: 0.0002575938997324556
step: 220, loss: 0.00016942177899181843
step: 230, loss: 0.000752425636164844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.987598647125141, f1=0.9808773903262092,
best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05306978151202202
step: 10, loss: 0.00041692337254062295
step: 20, loss: 0.0003588172548916191
step: 30, loss: 0.006871705874800682
step: 40, loss: 0.001126812887378037
step: 50, loss: 0.00021774145716335624
step: 60, loss: 0.0018260934157297015
step: 70, loss: 0.00016746365872677416
step: 80, loss: 0.00011361246288288385
step: 90, loss: 0.000314073811750859
step: 100, loss: 0.0009688932332210243
step: 110, loss: 0.0012229365529492497

```

```

step: 120, loss: 0.0039289300329983234
step: 130, loss: 0.0016439455794170499
step: 140, loss: 0.00012698063801508397
step: 150, loss: 0.00012894667452201247
step: 160, loss: 0.00011787872790591791
step: 170, loss: 0.0014194577233865857
step: 180, loss: 0.0002588476927485317
step: 190, loss: 0.0008869912126101553
step: 200, loss: 0.000693966168910265
step: 210, loss: 0.00013018480967730284
step: 220, loss: 0.0009023434249684215
step: 230, loss: 0.02561270259320736
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9876265466816648, f1=0.9776286353467561,
best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02614181488752365
step: 10, loss: 0.0003894486289937049
step: 20, loss: 0.0001691048964858055
step: 30, loss: 0.0001443415676476434
step: 40, loss: 9.452896483708173e-05
step: 50, loss: 0.003872159868478775
step: 60, loss: 8.34703678265214e-05
step: 70, loss: 0.00014245350030250847
step: 80, loss: 0.00023160944692790508
step: 90, loss: 8.847714343573898e-05
step: 100, loss: 0.00013198181113693863
step: 110, loss: 0.00010127199493581429
step: 120, loss: 7.907787949079648e-05
step: 130, loss: 0.0011901314137503505
step: 140, loss: 6.431037036236376e-05
step: 150, loss: 0.0003456837439443916
step: 160, loss: 0.0010111257433891296
step: 170, loss: 0.005187006201595068
step: 180, loss: 0.002494465559720993
step: 190, loss: 0.00017011504678521305
step: 200, loss: 0.008176170289516449
step: 210, loss: 0.00011096572416136041
step: 220, loss: 0.00011783481750171632
step: 230, loss: 0.02877018041908741
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9876265466816648, f1=0.9755011135857461,

```

```

best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002630485105328262
step: 10, loss: 0.00038106675492599607
step: 20, loss: 0.0004773887631017715
step: 30, loss: 0.00010769997606985271
step: 40, loss: 0.00038660771679133177
step: 50, loss: 0.00012037927081109956
step: 60, loss: 8.471996989101171e-05
step: 70, loss: 0.00010569590085651726
step: 80, loss: 0.03164425119757652
step: 90, loss: 0.003772253170609474
step: 100, loss: 8.268746751127765e-05
step: 110, loss: 0.004291401244699955
step: 120, loss: 0.03978458046913147
step: 130, loss: 8.019609231268987e-05
step: 140, loss: 8.827394776744768e-05
step: 150, loss: 5.7505469158058986e-05
step: 160, loss: 8.149105269694701e-05
step: 170, loss: 7.358257425948977e-05
step: 180, loss: 8.163518214132637e-05
step: 190, loss: 5.640167000819929e-05
step: 200, loss: 8.056374645093456e-05
step: 210, loss: 7.219695544335991e-05
step: 220, loss: 0.03520050644874573
step: 230, loss: 0.006526939570903778
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9875424688561721, f1=0.9817767653758542,
best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.5941087086685e-05
step: 10, loss: 5.287729436531663e-05
step: 20, loss: 5.0859096518252045e-05
step: 30, loss: 7.019426266197115e-05
step: 40, loss: 5.4034284403314814e-05
step: 50, loss: 0.0001479951461078599
step: 60, loss: 0.023608757182955742
step: 70, loss: 5.03816336276941e-05
step: 80, loss: 5.133826198289171e-05
step: 90, loss: 5.256481017568149e-05
step: 100, loss: 6.100775135564618e-05
step: 110, loss: 0.0014519451651722193

```

```

step: 120, loss: 0.018663935363292694
step: 130, loss: 0.11642929166555405
step: 140, loss: 0.04501138627529144
step: 150, loss: 6.387141911545768e-05
step: 160, loss: 0.0017798570916056633
step: 170, loss: 0.00010299465066054836
step: 180, loss: 0.0008764813537709415
step: 190, loss: 6.89322259859182e-05
step: 200, loss: 5.680702452082187e-05
step: 210, loss: 6.33980380371213e-05
step: 220, loss: 0.0005572677473537624
step: 230, loss: 7.901431672507897e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.987598647125141, f1=0.9776286353467561,
best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.703309918520972e-05
step: 10, loss: 0.0001285618491237983
step: 20, loss: 4.505127071752213e-05
step: 30, loss: 8.382656233152375e-05
step: 40, loss: 0.001654694089666009
step: 50, loss: 0.0015111872926354408
step: 60, loss: 7.236544479383156e-05
step: 70, loss: 7.354377157753333e-05
step: 80, loss: 0.00014844382531009614
step: 90, loss: 0.00018515653209760785
step: 100, loss: 7.862629718147218e-05
step: 110, loss: 5.3219107940094545e-05
step: 120, loss: 9.919055446516722e-05
step: 130, loss: 5.04778690810781e-05
step: 140, loss: 4.9024449253920466e-05
step: 150, loss: 3.840623321593739e-05
step: 160, loss: 0.00010132192983292043
step: 170, loss: 0.002949133049696684
step: 180, loss: 0.00028133162413723767
step: 190, loss: 5.891770342714153e-05
step: 200, loss: 8.111818897305056e-05
step: 210, loss: 4.520060247159563e-05
step: 220, loss: 5.3551048040390015e-05
step: 230, loss: 0.020706722512841225
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9875424688561721, f1=0.9806598407281,

```

```

best_f1=0.9788182831661093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.4128264309838414e-05
step: 10, loss: 0.02751884236931801
step: 20, loss: 4.759505100082606e-05
step: 30, loss: 0.0001589954918017611
step: 40, loss: 3.9806636777939275e-05
step: 50, loss: 3.7213958421489224e-05
step: 60, loss: 0.008165490813553333
step: 70, loss: 4.3013460526708513e-05
step: 80, loss: 3.4878674341598526e-05
step: 90, loss: 4.0841994632501155e-05
step: 100, loss: 3.0233717552619055e-05
step: 110, loss: 4.567385258269496e-05
step: 120, loss: 5.568888445850462e-05
step: 130, loss: 5.3224135626805946e-05
step: 140, loss: 6.264565308811143e-05
step: 150, loss: 5.219809463596903e-05
step: 160, loss: 0.02610957808792591
step: 170, loss: 5.550269634113647e-05
step: 180, loss: 3.8905163819435984e-05
step: 190, loss: 0.0001697386906016618
step: 200, loss: 0.009763195179402828
step: 210, loss: 3.782525163842365e-05
step: 220, loss: 0.0001568483276059851
step: 230, loss: 5.281061021378264e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.9898305084745763, f1=0.9775784753363228,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007552892784588039
step: 10, loss: 0.00011546217137947679
step: 20, loss: 0.0001909999264171347
step: 30, loss: 0.022006724029779434
step: 40, loss: 0.00037888268707320094
step: 50, loss: 5.9204794524703175e-05
step: 60, loss: 0.00011236253340030089
step: 70, loss: 0.0006680102669633925

```



```

step: 80, loss: 3.563093559932895e-05
step: 90, loss: 3.239788566133939e-05
step: 100, loss: 4.740253643831238e-05
step: 110, loss: 4.044766319566406e-05
step: 120, loss: 0.00017985621525440365
step: 130, loss: 5.3506766562350094e-05
step: 140, loss: 0.0001458782935515046
step: 150, loss: 0.025020619854331017
step: 160, loss: 0.010828288272023201
step: 170, loss: 4.903222361463122e-05
step: 180, loss: 0.0002662077604327351
step: 190, loss: 3.533683775458485e-05
step: 200, loss: 5.172009696252644e-05
step: 210, loss: 0.0001158702652901411
step: 220, loss: 2.7953903554589488e-05
step: 230, loss: 5.9898633480770513e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9875706214689265, f1=0.9831271091113611,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012097672879463062
step: 10, loss: 3.488158108666539e-05
step: 20, loss: 0.0042335172183811665
step: 30, loss: 6.217161717358977e-05
step: 40, loss: 3.838567863567732e-05
step: 50, loss: 6.284071423579007e-05
step: 60, loss: 5.0700968131422997e-05
step: 70, loss: 3.805913365795277e-05
step: 80, loss: 5.530457929125987e-05
step: 90, loss: 5.5357158998958766e-05
step: 100, loss: 0.004672108683735132
step: 110, loss: 4.8471163609065115e-05
step: 120, loss: 4.2751518776640296e-05
step: 130, loss: 5.5995958973653615e-05
step: 140, loss: 0.00012030195648549125
step: 150, loss: 0.00010790058149723336
step: 160, loss: 2.9216354960226454e-05
step: 170, loss: 3.1913365091895685e-05
step: 180, loss: 5.861303361598402e-05
step: 190, loss: 0.0018933399114757776
step: 200, loss: 3.5288496292196214e-05
step: 210, loss: 0.020517216995358467
step: 220, loss: 6.488867802545428e-05
step: 230, loss: 2.1595185899059288e-05

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9875706214689265, f1=0.9831271091113611,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.189353057881817e-05
step: 10, loss: 7.864476356189698e-05
step: 20, loss: 3.9519756683148444e-05
step: 30, loss: 0.00010247294994769618
step: 40, loss: 5.959613190498203e-05
step: 50, loss: 4.8184032493736595e-05
step: 60, loss: 2.7033671358367428e-05
step: 70, loss: 4.413514398038387e-05
step: 80, loss: 0.00023746976512484252
step: 90, loss: 2.7565849450184032e-05
step: 100, loss: 4.310340227675624e-05
step: 110, loss: 4.692485890700482e-05
step: 120, loss: 8.521191193722188e-05
step: 130, loss: 4.859680848312564e-05
step: 140, loss: 0.0001708974305074662
step: 150, loss: 5.3532708989223465e-05
step: 160, loss: 0.0012795928632840514
step: 170, loss: 2.6735675419331528e-05
step: 180, loss: 3.881557495333254e-05
step: 190, loss: 3.707639916683547e-05
step: 200, loss: 3.8286852941382676e-05
step: 210, loss: 0.0026611117646098137
step: 220, loss: 0.00011211300443392247
step: 230, loss: 5.4937281674938276e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9875706214689265, f1=0.9820224719101124,
best_f1=0.9775784753363228
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True

```

```

loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 241.29it/s]
load_f1 = 0.9886877828054299
real_f1 = 0.9898305084745763
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:16, 259.76it/s]

```

### 1.1.3 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--da entry_swap \
--da span_shuffle \
--summarize

```

```
# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt
```

```
# Step 3: run the matcher
```

```
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Downloading: 100% 442/442 [00:00<00:00, 378kB/s]

Downloading: 100% 232k/232k [00:00<00:00, 257kB/s]

Downloading: 100% 268M/268M [00:03<00:00, 69.5MB/s]

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>",  
UserWarning)

step: 0, loss: 0.7851431369781494

step: 10, loss: 0.4201892912387848

step: 20, loss: 0.49773260951042175

step: 30, loss: 0.40549901127815247

step: 40, loss: 0.33648571372032166

step: 50, loss: 0.19393065571784973

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 60, loss: 0.2025073617696762

step: 70, loss: 0.05367906019091606

step: 80, loss: 0.17749850451946259

step: 90, loss: 0.28770211338996887

step: 100, loss: 0.2695692181587219

step: 110, loss: 0.09074883908033371

step: 120, loss: 0.05067526549100876

step: 130, loss: 0.04895973950624466

step: 140, loss: 0.22394263744354248

step: 150, loss: 0.03799163177609444

step: 160, loss: 0.12455091625452042

step: 170, loss: 0.15918491780757904

step: 180, loss: 0.12400630116462708

step: 190, loss: 0.04283490777015686

step: 200, loss: 0.16901208460330963

step: 210, loss: 0.07958915084600449

step: 220, loss: 0.1124597042798996

step: 230, loss: 0.1223619356751442

step: 240, loss: 0.07758272439241409

step: 250, loss: 0.11241567134857178

step: 260, loss: 0.0192920733243227

step: 270, loss: 0.012755652889609337

step: 280, loss: 0.10133209079504013

step: 290, loss: 0.08258091658353806

step: 300, loss: 0.09657403081655502

step: 310, loss: 0.06266164779663086

step: 320, loss: 0.07034933567047119

step: 330, loss: 0.1020229309797287

step: 340, loss: 0.1332031935453415

step: 350, loss: 0.07131805270910263

step: 360, loss: 0.08827780932188034

step: 370, loss: 0.15301422774791718

step: 380, loss: 0.18991851806640625

step: 390, loss: 0.01886887662112713

step: 400, loss: 0.012111061252653599

```

step: 410, loss: 0.023569846525788307
step: 420, loss: 0.018679434433579445
step: 430, loss: 0.07483624666929245
step: 440, loss: 0.06355506926774979
step: 450, loss: 0.011745438911020756
step: 460, loss: 0.11229513585567474
step: 470, loss: 0.2384861558675766
step: 480, loss: 0.3367914855480194
step: 490, loss: 0.03437238559126854
step: 500, loss: 0.03362399712204933
step: 510, loss: 0.06350122392177582
step: 520, loss: 0.04784662649035454
step: 530, loss: 0.07806862145662308
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9341983317886932, f1=0.9264229523368811,
best_f1=0.9264229523368811
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09310124069452286
step: 10, loss: 0.1258852630853653
step: 20, loss: 0.17387156188488007
step: 30, loss: 0.09424588829278946
step: 40, loss: 0.010005155578255653
step: 50, loss: 0.09569816291332245
step: 60, loss: 0.18408967554569244
step: 70, loss: 0.09453635662794113
step: 80, loss: 0.015619453974068165
step: 90, loss: 0.006738041527569294
step: 100, loss: 0.2365001142024994
step: 110, loss: 0.012322367168962955
step: 120, loss: 0.046765413135290146
step: 130, loss: 0.023075882345438004
step: 140, loss: 0.012116449885070324
step: 150, loss: 0.0498468279838562
step: 160, loss: 0.04486916586756706
step: 170, loss: 0.07949409633874893
step: 180, loss: 0.015400931239128113
step: 190, loss: 0.06609617173671722
step: 200, loss: 0.01714053377509117
step: 210, loss: 0.008827749639749527
step: 220, loss: 0.16438624262809753

```

```

step: 230, loss: 0.029715409502387047
step: 240, loss: 0.13579480350017548
step: 250, loss: 0.024169480428099632
step: 260, loss: 0.010707014240324497
step: 270, loss: 0.04980507865548134
step: 280, loss: 0.1769968718290329
step: 290, loss: 0.07462439686059952
step: 300, loss: 0.042237039655447006
step: 310, loss: 0.1054031178355217
step: 320, loss: 0.20289120078086853
step: 330, loss: 0.05856379121541977
step: 340, loss: 0.043319035321474075
step: 350, loss: 0.03493563085794449
step: 360, loss: 0.047485772520303726
step: 370, loss: 0.005797842051833868
step: 380, loss: 0.05421380326151848
step: 390, loss: 0.04036106914281845
step: 400, loss: 0.035100217908620834
step: 410, loss: 0.0004353119875304401
step: 420, loss: 0.07176195830106735
step: 430, loss: 0.02526124194264412
step: 440, loss: 0.009141308255493641
step: 450, loss: 0.02919033356010914
step: 460, loss: 0.21513374149799347
step: 470, loss: 0.11248403042554855
step: 480, loss: 0.2316686511039734
step: 490, loss: 0.03672170266509056
step: 500, loss: 0.01599409431219101
step: 510, loss: 0.06361118704080582
step: 520, loss: 0.05857358127832413
step: 530, loss: 0.12405149638652802
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9377859103385179, f1=0.9302961275626425,
best_f1=0.9302961275626425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03214334696531296
step: 10, loss: 0.057088661938905716
step: 20, loss: 0.16988280415534973
step: 30, loss: 0.2740074396133423

```

step: 40, loss: 0.005195838399231434  
step: 50, loss: 0.024433165788650513  
step: 60, loss: 0.005310152657330036  
step: 70, loss: 0.024092311039566994  
step: 80, loss: 0.0029906178824603558  
step: 90, loss: 0.017286475747823715  
step: 100, loss: 0.16029752790927887  
step: 110, loss: 0.006428631022572517  
step: 120, loss: 0.007265985943377018  
step: 130, loss: 0.017994899302721024  
step: 140, loss: 0.02498358115553856  
step: 150, loss: 0.005884350277483463  
step: 160, loss: 0.0177391916513443  
step: 170, loss: 0.006762017961591482  
step: 180, loss: 0.03615398332476616  
step: 190, loss: 0.04871947318315506  
step: 200, loss: 0.009070922620594501  
step: 210, loss: 0.11827580630779266  
step: 220, loss: 0.019596947357058525  
step: 230, loss: 0.03234519064426422  
step: 240, loss: 0.006056963466107845  
step: 250, loss: 0.0037725670263171196  
step: 260, loss: 0.0018008783226832747  
step: 270, loss: 0.013197176158428192  
step: 280, loss: 0.010041642002761364  
step: 290, loss: 0.01768609881401062  
step: 300, loss: 0.2394009679555893  
step: 310, loss: 0.09940727800130844  
step: 320, loss: 0.13719448447227478  
step: 330, loss: 0.0052350424230098724  
step: 340, loss: 0.002134852111339569  
step: 350, loss: 0.013275153934955597  
step: 360, loss: 0.020102713257074356  
step: 370, loss: 0.002202104777097702  
step: 380, loss: 0.021342653781175613  
step: 390, loss: 0.004707434214651585  
step: 400, loss: 0.008919449523091316  
step: 410, loss: 0.02072722837328911  
step: 420, loss: 0.04720411077141762  
step: 430, loss: 0.004297892563045025  
step: 440, loss: 0.0332987904548645  
step: 450, loss: 0.03112950176000595  
step: 460, loss: 0.055095914751291275  
step: 470, loss: 0.0033411895856261253  
step: 480, loss: 0.0018320673843845725  
step: 490, loss: 0.01894492097198963  
step: 500, loss: 0.10346455127000809  
step: 510, loss: 0.004490478429943323



```

step: 520, loss: 0.02006426826119423
step: 530, loss: 0.017355887219309807
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9324515824279641, f1=0.9324577861163227,
best_f1=0.9302961275626425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022364437580108643
step: 10, loss: 0.008457357063889503
step: 20, loss: 0.011430800892412663
step: 30, loss: 0.005267960485070944
step: 40, loss: 0.0030379092786461115
step: 50, loss: 0.0038230018690228462
step: 60, loss: 0.0065214610658586025
step: 70, loss: 0.0007932856678962708
step: 80, loss: 0.053223490715026855
step: 90, loss: 0.02715209312736988
step: 100, loss: 0.1019861027598381
step: 110, loss: 0.0011034533381462097
step: 120, loss: 0.001831911620683968
step: 130, loss: 0.03350096568465233
step: 140, loss: 0.0270525012165308
step: 150, loss: 0.0007984975236468017
step: 160, loss: 0.01619395613670349
step: 170, loss: 0.0007885352242738008
step: 180, loss: 0.0065422565676271915
step: 190, loss: 0.010833630338311195
step: 200, loss: 0.0012308793375268579
step: 210, loss: 0.20430438220500946
step: 220, loss: 0.018287522718310356
step: 230, loss: 0.1972569078207016
step: 240, loss: 0.007958633825182915
step: 250, loss: 0.001272503868676722
step: 260, loss: 0.09367763251066208
step: 270, loss: 0.02285863645374775
step: 280, loss: 0.0028575737960636616
step: 290, loss: 0.13022686541080475
step: 300, loss: 0.0026353870052844286
step: 310, loss: 0.0011080035474151373
step: 320, loss: 0.0915035754442215
step: 330, loss: 0.19580644369125366
step: 340, loss: 0.21421480178833008
step: 350, loss: 0.00976715236902237
step: 360, loss: 0.004942188039422035
step: 370, loss: 0.21076349914073944

```

```

step: 380, loss: 0.013198225758969784
step: 390, loss: 0.04073679447174072
step: 400, loss: 0.009459023363888264
step: 410, loss: 0.027208592742681503
step: 420, loss: 0.0016756131080910563
step: 430, loss: 0.008654954843223095
step: 440, loss: 0.06490220129489899
step: 450, loss: 0.055550768971443176
step: 460, loss: 0.004378024023026228
step: 470, loss: 0.008770917542278767
step: 480, loss: 0.005013924557715654
step: 490, loss: 0.05985459312796593
step: 500, loss: 0.012657653540372849
step: 510, loss: 0.012625527568161488
step: 520, loss: 0.003087430028244853
step: 530, loss: 0.007818463258445263
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9350893697083725, f1=0.935258500232883,
best_f1=0.9302961275626425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003724132664501667
step: 10, loss: 0.007625632453709841
step: 20, loss: 0.0007697132532484829
step: 30, loss: 0.002743219491094351
step: 40, loss: 0.03342575207352638
step: 50, loss: 0.023823430761694908
step: 60, loss: 0.005213251803070307
step: 70, loss: 0.00035469321301206946
step: 80, loss: 0.0007852651178836823
step: 90, loss: 0.020740818232297897
step: 100, loss: 0.0023057940416038036
step: 110, loss: 0.007025258149951696
step: 120, loss: 0.005268969107419252
step: 130, loss: 0.004030435346066952
step: 140, loss: 0.0013091827277094126
step: 150, loss: 0.00127311737742275
step: 160, loss: 0.14310070872306824
step: 170, loss: 0.027747545391321182
step: 180, loss: 0.004836830776184797
step: 190, loss: 0.0016158653888851404
step: 200, loss: 0.0004110459121875465
step: 210, loss: 0.016447534784674644
step: 220, loss: 0.0001634812942938879
step: 230, loss: 0.001377974869683385

```

```

step: 240, loss: 0.010198652744293213
step: 250, loss: 0.00025871649268083274
step: 260, loss: 0.0005622024182230234
step: 270, loss: 0.006556901149451733
step: 280, loss: 0.11876976490020752
step: 290, loss: 0.0006740431999787688
step: 300, loss: 0.007985607720911503
step: 310, loss: 0.05467064678668976
step: 320, loss: 0.01099750678986311
step: 330, loss: 0.0014033896150067449
step: 340, loss: 0.011604811064898968
step: 350, loss: 0.0020165536552667618
step: 360, loss: 0.001142143621109426
step: 370, loss: 0.00038431177381426096
step: 380, loss: 0.016141749918460846
step: 390, loss: 0.016289642080664635
step: 400, loss: 0.08760735392570496
step: 410, loss: 0.0007343165343627334
step: 420, loss: 0.00420590303838253
step: 430, loss: 0.03819920867681503
step: 440, loss: 0.0026575042866170406
step: 450, loss: 0.00685600470751524
step: 460, loss: 0.0037089113611727953
step: 470, loss: 0.06599760055541992
step: 480, loss: 0.0029625562019646168
step: 490, loss: 0.009303403086960316
step: 500, loss: 0.0073508634231984615
step: 510, loss: 0.0017167876940220594
step: 520, loss: 0.0011558301048353314
step: 530, loss: 0.005571433808654547
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9421564090698751, f1=0.9348623853211009,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003672955324873328
step: 10, loss: 0.0008555946405977011
step: 20, loss: 0.00018110897508449852
step: 30, loss: 0.002166388789191842
step: 40, loss: 0.0034356729593127966
step: 50, loss: 0.0002713593712542206

```

step: 60, loss: 0.0010615445207804441  
step: 70, loss: 0.006212301552295685  
step: 80, loss: 0.0009519605664536357  
step: 90, loss: 0.011276218108832836  
step: 100, loss: 0.13839617371559143  
step: 110, loss: 0.01847187802195549  
step: 120, loss: 0.000880268809851259  
step: 130, loss: 0.00044087396236136556  
step: 140, loss: 0.000735597568564117  
step: 150, loss: 0.0014448907459154725  
step: 160, loss: 0.00032389882835559547  
step: 170, loss: 0.00022215876379050314  
step: 180, loss: 0.0005211875541135669  
step: 190, loss: 0.005227343179285526  
step: 200, loss: 0.000458481750683859  
step: 210, loss: 0.0004724983300548047  
step: 220, loss: 0.00017403101082891226  
step: 230, loss: 0.00014266608923207968  
step: 240, loss: 0.0005096546956337988  
step: 250, loss: 0.00022320142306853086  
step: 260, loss: 0.004673474933952093  
step: 270, loss: 0.0027819450479000807  
step: 280, loss: 0.0011166853364557028  
step: 290, loss: 0.0006278416840359569  
step: 300, loss: 0.0013944994425401092  
step: 310, loss: 0.014375299215316772  
step: 320, loss: 0.025089845061302185  
step: 330, loss: 0.0009489263757131994  
step: 340, loss: 0.03469469025731087  
step: 350, loss: 0.046246834099292755  
step: 360, loss: 0.005997939966619015  
step: 370, loss: 0.018211163580417633  
step: 380, loss: 0.0009585131192579865  
step: 390, loss: 0.1267774999141693  
step: 400, loss: 0.0006177115719765425  
step: 410, loss: 0.00021643601940013468  
step: 420, loss: 0.0026310873217880726  
step: 430, loss: 0.0009529827511869371  
step: 440, loss: 0.03838116303086281  
step: 450, loss: 0.010829092934727669  
step: 460, loss: 0.03043299913406372  
step: 470, loss: 0.03015531785786152  
step: 480, loss: 0.007127454504370689  
step: 490, loss: 0.00017339395708404481  
step: 500, loss: 0.0010574538027867675  
step: 510, loss: 0.0003461516462266445  
step: 520, loss: 0.000978580443188548  
step: 530, loss: 0.032975222915410995

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9345622119815669, f1=0.935528120713306,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019160045310854912
step: 10, loss: 0.010273144580423832
step: 20, loss: 0.12015868723392487
step: 30, loss: 0.0019343476742506027
step: 40, loss: 0.00011104735312983394
step: 50, loss: 0.0015734587796032429
step: 60, loss: 0.1175215020775795
step: 70, loss: 0.006226715631783009
step: 80, loss: 0.0003986794035881758
step: 90, loss: 0.0003851470828521997
step: 100, loss: 0.0004567478608805686
step: 110, loss: 0.02296859212219715
step: 120, loss: 0.028982367366552353
step: 130, loss: 0.006452678702771664
step: 140, loss: 0.01796760782599449
step: 150, loss: 0.00010197139636147767
step: 160, loss: 0.00028826549532823265
step: 170, loss: 0.00034266512375324965
step: 180, loss: 0.0004931723233312368
step: 190, loss: 9.805302397580817e-05
step: 200, loss: 0.014863875694572926
step: 210, loss: 0.00012736600183416158
step: 220, loss: 0.00015618845645803958
step: 230, loss: 0.0014113321667537093
step: 240, loss: 0.0005745814414694905
step: 250, loss: 0.007699222769588232
step: 260, loss: 0.04662683606147766
step: 270, loss: 8.391316805500537e-05
step: 280, loss: 0.08533324301242828
step: 290, loss: 0.006867149844765663
step: 300, loss: 0.009244895540177822
step: 310, loss: 0.0007939627976156771
step: 320, loss: 0.05883963778614998
step: 330, loss: 0.00019247901218477637
step: 340, loss: 0.003741280408576131
step: 350, loss: 0.0038005667738616467
step: 360, loss: 0.004984613042324781
step: 370, loss: 0.0001729572977637872
step: 380, loss: 0.0013504928210750222
step: 390, loss: 0.01128759328275919
```

```

step: 400, loss: 8.160275319823995e-05
step: 410, loss: 0.0019195991335436702
step: 420, loss: 0.0002421910030534491
step: 430, loss: 0.0001965109258890152
step: 440, loss: 0.0007646969170309603
step: 450, loss: 0.0014205614570528269
step: 460, loss: 0.003480158979073167
step: 470, loss: 0.03479956090450287
step: 480, loss: 0.008760296739637852
step: 490, loss: 0.0008904307032935321
step: 500, loss: 0.00013753483653999865
step: 510, loss: 0.005785753484815359
step: 520, loss: 0.0016420171596109867
step: 530, loss: 0.0011497506638988853
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9364858599907279, f1=0.9376718606782769,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014097001403570175
step: 10, loss: 0.014324956573545933
step: 20, loss: 0.023374728858470917
step: 30, loss: 0.015417173504829407
step: 40, loss: 0.00012069562944816425
step: 50, loss: 0.003828223329037428
step: 60, loss: 0.0005546334432438016
step: 70, loss: 0.0018337168730795383
step: 80, loss: 0.0001988643780350685
step: 90, loss: 0.003188646398484707
step: 100, loss: 0.0003817038668785244
step: 110, loss: 0.0005009302403777838
step: 120, loss: 0.002960646292194724
step: 130, loss: 0.0012737575452774763
step: 140, loss: 0.00014597202243749052
step: 150, loss: 0.0002035445795627311
step: 160, loss: 0.0006779509712941945
step: 170, loss: 0.001474561751820147
step: 180, loss: 0.00011804235691670328
step: 190, loss: 0.0010402232874184847
step: 200, loss: 0.007770396303385496
step: 210, loss: 0.017354043200612068
step: 220, loss: 0.001763945329003036
step: 230, loss: 0.0002818196371663362
step: 240, loss: 6.297280197031796e-05
step: 250, loss: 0.00010616119106998667

```

```

step: 260, loss: 0.017266076058149338
step: 270, loss: 4.2390911403344944e-05
step: 280, loss: 0.009047284722328186
step: 290, loss: 0.003087734105065465
step: 300, loss: 0.00025652110343798995
step: 310, loss: 0.030011283233761787
step: 320, loss: 9.302244143327698e-05
step: 330, loss: 0.011541013605892658
step: 340, loss: 0.0003923702461179346
step: 350, loss: 0.017809387296438217
step: 360, loss: 0.0035493001341819763
step: 370, loss: 0.0011405774857848883
step: 380, loss: 0.0006059047882445157
step: 390, loss: 0.01569078117609024
step: 400, loss: 0.08777042478322983
step: 410, loss: 0.0015423534205183387
step: 420, loss: 0.0001678274420555681
step: 430, loss: 0.00021183221542742103
step: 440, loss: 0.00015710234583821148
step: 450, loss: 0.000846532522700727
step: 460, loss: 0.0019513519946485758
step: 470, loss: 0.00018546673527453095
step: 480, loss: 0.0007536725606769323
step: 490, loss: 0.0012748728040605783
step: 500, loss: 0.0008528146427124739
step: 510, loss: 0.0013829240342602134
step: 520, loss: 0.002762907650321722
step: 530, loss: 0.010340556502342224
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9386814200092208, f1=0.9365586490187129,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.924886788008735e-05
step: 10, loss: 7.443966751452535e-05
step: 20, loss: 0.0007529522408731282
step: 30, loss: 9.37742879614234e-05
step: 40, loss: 0.002352998126298189
step: 50, loss: 0.00011768287367885932
step: 60, loss: 0.00015113138942979276
step: 70, loss: 0.12077303975820541
step: 80, loss: 0.00014646680210717022
step: 90, loss: 0.0008440185338258743
step: 100, loss: 0.0001221085840370506
step: 110, loss: 0.0004334732366260141

```

```
step: 120, loss: 0.00014336516323965043
step: 130, loss: 0.0001747872302075848
step: 140, loss: 0.00036358466604724526
step: 150, loss: 0.002845622133463621
step: 160, loss: 0.00019482389325276017
step: 170, loss: 5.262591002974659e-05
step: 180, loss: 0.005560826975852251
step: 190, loss: 0.0002584619796834886
step: 200, loss: 0.001070503261871636
step: 210, loss: 0.00010046445822808892
step: 220, loss: 8.973978401627392e-05
step: 230, loss: 0.00020483601838350296
step: 240, loss: 5.770728239440359e-05
step: 250, loss: 0.0008312372374348342
step: 260, loss: 0.00020137312822043896
step: 270, loss: 0.0004414245777297765
step: 280, loss: 0.0004065689572598785
step: 290, loss: 8.408827125094831e-05
step: 300, loss: 0.00012527662329375744
step: 310, loss: 4.6437762648565695e-05
step: 320, loss: 0.0018772230250760913
step: 330, loss: 3.90533241443336e-05
step: 340, loss: 6.285660492721945e-05
step: 350, loss: 8.00335910753347e-05
step: 360, loss: 0.03267737478017807
step: 370, loss: 6.752402987331152e-05
step: 380, loss: 5.1572911615949124e-05
step: 390, loss: 0.001954442821443081
step: 400, loss: 0.008506237529218197
step: 410, loss: 0.0005186077323742211
step: 420, loss: 0.0017029957380145788
step: 430, loss: 3.672597085824236e-05
step: 440, loss: 8.31182042020373e-05
step: 450, loss: 0.0002015769568970427
step: 460, loss: 0.00021048109920229763
step: 470, loss: 0.0010747263440862298
step: 480, loss: 0.0014969143085181713
step: 490, loss: 4.918070771964267e-05
step: 500, loss: 0.057752493768930435
step: 510, loss: 0.00015612943388987333
step: 520, loss: 0.024129383265972137
step: 530, loss: 4.498753332882188e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9409048938134812, f1=0.9405306495882891,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```



An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

step: 0, loss: 0.025972425937652588  
step: 10, loss: 2.367345405218657e-05  
step: 20, loss: 7.868710963521153e-05  
step: 30, loss: 0.004152718000113964  
step: 40, loss: 0.00013493614096660167  
step: 50, loss: 4.3928954255534336e-05  
step: 60, loss: 6.685643893433735e-05  
step: 70, loss: 0.20477241277694702  
step: 80, loss: 0.0037024449557065964  
step: 90, loss: 9.253049211111092e-05  
step: 100, loss: 6.488831422757357e-05  
step: 110, loss: 0.0009255376062355936  
step: 120, loss: 7.435529551003128e-05  
step: 130, loss: 6.708482396788895e-05  
step: 140, loss: 9.18396472116001e-05  
step: 150, loss: 6.318526720860973e-05  
step: 160, loss: 0.0007045388920232654  
step: 170, loss: 3.219606151105836e-05  
step: 180, loss: 0.0005692818085663021  
step: 190, loss: 0.008012933656573296  
step: 200, loss: 0.0003460860170889646  
step: 210, loss: 7.037873001536354e-05  
step: 220, loss: 0.0009844249580055475  
step: 230, loss: 0.00013099770876578987  
step: 240, loss: 4.6178960474208e-05  
step: 250, loss: 0.0002202346222475171  
step: 260, loss: 0.004300322383642197  
step: 270, loss: 0.0001890730782179162  
step: 280, loss: 5.7717184972716495e-05  
step: 290, loss: 0.0008702754857949913  
step: 300, loss: 0.007907029241323471  
step: 310, loss: 0.02358667366206646  
step: 320, loss: 0.0006441404693759978  
step: 330, loss: 4.20035285060294e-05  
step: 340, loss: 0.000666503852698952  
step: 350, loss: 5.1153885578969494e-05  
step: 360, loss: 0.000576186750549823  
step: 370, loss: 0.06933163106441498  
step: 380, loss: 0.00020298083836678416  
step: 390, loss: 7.611681940034032e-05  
step: 400, loss: 0.00023874379985500127  
step: 410, loss: 0.012313381768763065  
step: 420, loss: 0.02363990806043148  
step: 430, loss: 0.00018967717187479138  
step: 440, loss: 4.594210258801468e-05  
step: 450, loss: 0.004388879518955946

```

step: 460, loss: 0.0012651998549699783
step: 470, loss: 0.00010492288129171357
step: 480, loss: 0.010652796365320683
step: 490, loss: 0.0508611835539341
step: 500, loss: 0.010589934885501862
step: 510, loss: 0.0005469421739690006
step: 520, loss: 0.0034139957278966904
step: 530, loss: 2.6296169380657375e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9320297951582867, f1=0.9357967667436491,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020220858976244926
step: 10, loss: 0.035214681178331375
step: 20, loss: 0.00046241015661507845
step: 30, loss: 0.04596128687262535
step: 40, loss: 0.0002853407640941441
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.001547411666251719
step: 60, loss: 9.948639490175992e-05
step: 70, loss: 0.006370405200868845
step: 80, loss: 0.00012410327326506376
step: 90, loss: 0.001123660011216998
step: 100, loss: 0.006866804789751768
step: 110, loss: 4.7340061428258196e-05
step: 120, loss: 0.0005065469304099679
step: 130, loss: 0.0002070733462460339
step: 140, loss: 2.2503976651933044e-05
step: 150, loss: 0.0007848814129829407
step: 160, loss: 0.0006114402785897255
step: 170, loss: 0.0001965171832125634
step: 180, loss: 3.332869164296426e-05
step: 190, loss: 0.0004349601804278791
step: 200, loss: 0.0010017547756433487
step: 210, loss: 9.2340684204828e-05
step: 220, loss: 8.268686360679567e-05
step: 230, loss: 2.9413891752483323e-05
step: 240, loss: 3.635603570728563e-05
step: 250, loss: 0.00021864153677597642
step: 260, loss: 3.344007200212218e-05
step: 270, loss: 0.0012837451649829745
step: 280, loss: 8.327844261657447e-05
step: 290, loss: 0.00015781029651407152
step: 300, loss: 0.022363383322954178

```

```

step: 310, loss: 0.00020546979794744402
step: 320, loss: 0.019605128094553947
step: 330, loss: 0.0015704454854130745
step: 340, loss: 5.9590020100586116e-05
step: 350, loss: 0.0015087450155988336
step: 360, loss: 0.00013082334771752357
step: 370, loss: 3.651528095360845e-05
step: 380, loss: 3.86866377084516e-05
step: 390, loss: 0.0005798316560685635
step: 400, loss: 2.6403731681057252e-05
step: 410, loss: 0.0024193967692553997
step: 420, loss: 6.375926022883505e-05
step: 430, loss: 0.0004110498703084886
step: 440, loss: 2.951825808850117e-05
step: 450, loss: 0.0014104351866990328
step: 460, loss: 0.002569804899394512
step: 470, loss: 0.010712586343288422
step: 480, loss: 3.4404707548674196e-05
step: 490, loss: 0.11057578772306442
step: 500, loss: 0.00012430995411705226
step: 510, loss: 8.790374704403803e-05
step: 520, loss: 2.1851783458259888e-05
step: 530, loss: 2.3654931283090264e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9355432780847146, f1=0.9353369763205829,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.318813732126728e-05
step: 10, loss: 0.005317663308233023
step: 20, loss: 2.4082977688522078e-05
step: 30, loss: 3.5168795875506476e-05
step: 40, loss: 0.00010619130625855178
step: 50, loss: 0.007562640123069286
step: 60, loss: 0.0008691660477779806
step: 70, loss: 0.0002104428713209927
step: 80, loss: 0.0015309937298297882
step: 90, loss: 0.00013015992590226233
step: 100, loss: 4.1725168557604775e-05
step: 110, loss: 2.5740600904100575e-05
step: 120, loss: 0.0002256111183669418
step: 130, loss: 4.7247358452295884e-05
step: 140, loss: 3.080330498050898e-05
step: 150, loss: 0.01642211712896824
step: 160, loss: 3.355587614350952e-05

```

```

step: 170, loss: 0.0004045868699904531
step: 180, loss: 5.10653990204446e-05
step: 190, loss: 2.2880152755533345e-05
step: 200, loss: 4.489025741349906e-05
step: 210, loss: 0.00023921173124108464
step: 220, loss: 2.3334538127528504e-05
step: 230, loss: 0.0014024324482306838
step: 240, loss: 0.0015446798643097281
step: 250, loss: 0.00012788560707122087
step: 260, loss: 0.00011227862705709413
step: 270, loss: 2.8902011763420887e-05
step: 280, loss: 3.66342173947487e-05
step: 290, loss: 0.003644846845418215
step: 300, loss: 0.0013754760148003697
step: 310, loss: 4.066588371642865e-05
step: 320, loss: 2.7804831916000694e-05
step: 330, loss: 2.1554069462581538e-05
step: 340, loss: 9.297998622059822e-05
step: 350, loss: 0.013858577236533165
step: 360, loss: 0.0010337947169318795
step: 370, loss: 0.00021155434660613537
step: 380, loss: 0.00022434559650719166
step: 390, loss: 0.00034726824378594756
step: 400, loss: 2.4783075787127018e-05
step: 410, loss: 0.006674991920590401
step: 420, loss: 0.0010105771943926811
step: 430, loss: 8.753233851166442e-05
step: 440, loss: 0.001233719172887504
step: 450, loss: 3.760783511097543e-05
step: 460, loss: 0.009435749612748623
step: 470, loss: 3.5307559301145375e-05
step: 480, loss: 4.547188655124046e-05
step: 490, loss: 0.0024324501864612103
step: 500, loss: 0.0014602404553443193
step: 510, loss: 0.00014148584159556776
step: 520, loss: 0.031133972108364105
step: 530, loss: 5.1201117457821965e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9340101522842639, f1=0.9333333333333333,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003203652158845216
step: 10, loss: 0.0001646214077482
step: 20, loss: 0.029451392590999603

```

step: 30, loss: 5.257242446532473e-05  
step: 40, loss: 0.00022720941342413425  
step: 50, loss: 7.260640268214047e-05  
step: 60, loss: 0.00036530743818730116  
step: 70, loss: 3.124155045952648e-05  
step: 80, loss: 2.0224137188051827e-05  
step: 90, loss: 0.05903752148151398  
step: 100, loss: 0.0004910507705062628  
step: 110, loss: 8.596728002885357e-05  
step: 120, loss: 0.0007050620624795556  
step: 130, loss: 0.0015540359308943152  
step: 140, loss: 2.599343679321464e-05  
step: 150, loss: 8.067581802606583e-05  
step: 160, loss: 0.00020970423065591604  
step: 170, loss: 3.890359221259132e-05  
step: 180, loss: 1.9117815099889413e-05  
step: 190, loss: 0.00012252079613972455  
step: 200, loss: 0.002140800468623638  
step: 210, loss: 0.0010418411111459136  
step: 220, loss: 2.5785016987356357e-05  
step: 230, loss: 1.8849512343876995e-05  
step: 240, loss: 3.956046930397861e-05  
step: 250, loss: 0.04689835384488106  
step: 260, loss: 0.00033915278618223965  
step: 270, loss: 0.000133013934828341  
step: 280, loss: 0.00015488523058593273  
step: 290, loss: 3.3562679163878784e-05  
step: 300, loss: 4.090809670742601e-05  
step: 310, loss: 4.2450803448446095e-05  
step: 320, loss: 2.5245346478186548e-05  
step: 330, loss: 0.00014287943486124277  
step: 340, loss: 2.2444386559072882e-05  
step: 350, loss: 0.04925762116909027  
step: 360, loss: 0.0005209283553995192  
step: 370, loss: 0.00022384285693988204  
step: 380, loss: 0.00027071096701547503  
step: 390, loss: 4.4850479753222317e-05  
step: 400, loss: 0.0003064764605369419  
step: 410, loss: 5.787882400909439e-05  
step: 420, loss: 4.137981159146875e-05  
step: 430, loss: 0.0023018685169517994  
step: 440, loss: 3.0895782401785254e-05  
step: 450, loss: 3.091101461905055e-05  
step: 460, loss: 2.305081216036342e-05  
step: 470, loss: 0.03311459720134735  
step: 480, loss: 0.0008371368166990578  
step: 490, loss: 0.0005741185159422457  
step: 500, loss: 9.707412391435355e-05

```

step: 510, loss: 3.290753375040367e-05
step: 520, loss: 3.1289717298932374e-05
step: 530, loss: 8.046864968491718e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9366852886405959, f1=0.934313275149288,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011368022387614474
step: 10, loss: 0.00018710026051849127
step: 20, loss: 1.7139736883109435e-05
step: 30, loss: 5.908628008910455e-05
step: 40, loss: 0.0012599240290001035
step: 50, loss: 6.915505946381018e-05
step: 60, loss: 2.3312148186960258e-05
step: 70, loss: 1.6864085409906693e-05
step: 80, loss: 0.00012658516061492264
step: 90, loss: 1.9829152734018862e-05
step: 100, loss: 3.1741652492200956e-05
step: 110, loss: 2.2086609533289447e-05
step: 120, loss: 6.416637188522145e-05
step: 130, loss: 4.614689532900229e-05
step: 140, loss: 0.004518907051533461
step: 150, loss: 5.0516238843556494e-05
step: 160, loss: 0.0007346902857534587
step: 170, loss: 0.0006564934155903757
step: 180, loss: 0.00029146054293960333
step: 190, loss: 0.00016041500202845782
step: 200, loss: 9.631463035475463e-05
step: 210, loss: 6.925896013854071e-05
step: 220, loss: 0.0005382204544730484
step: 230, loss: 0.0012852057116106153
step: 240, loss: 0.0004994238261133432
step: 250, loss: 2.7354028134141117e-05
step: 260, loss: 1.7716949514579028e-05
step: 270, loss: 0.004102763254195452
step: 280, loss: 1.82870717253536e-05
step: 290, loss: 4.269152123015374e-05
step: 300, loss: 3.299823583802208e-05
step: 310, loss: 8.21104840724729e-05
step: 320, loss: 0.00018424454901833087
step: 330, loss: 2.5784733225009404e-05
step: 340, loss: 3.587646278901957e-05
step: 350, loss: 0.0015375545481219888
step: 360, loss: 0.003646369092166424

```

```

step: 370, loss: 9.016845433507115e-05
step: 380, loss: 4.196385998511687e-05
step: 390, loss: 0.000832017045468092
step: 400, loss: 2.266049887111876e-05
step: 410, loss: 1.5973819245118648e-05
step: 420, loss: 2.369964749959763e-05
step: 430, loss: 2.7510284780873917e-05
step: 440, loss: 1.922916817420628e-05
step: 450, loss: 3.690293669933453e-05
step: 460, loss: 0.000682301411870867
step: 470, loss: 2.5602888854336925e-05
step: 480, loss: 2.9606075258925557e-05
step: 490, loss: 5.76923084736336e-05
step: 500, loss: 0.00015076746058184654
step: 510, loss: 6.033023237250745e-05
step: 520, loss: 2.634082193253562e-05
step: 530, loss: 0.0004528408171609044
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9377593360995851, f1=0.9358215748748293,
best_f1=0.9348623853211009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.91247046738863e-05
step: 10, loss: 3.446843402343802e-05
step: 20, loss: 2.9417646146612242e-05
step: 30, loss: 3.879853102262132e-05
step: 40, loss: 0.0016101505607366562
step: 50, loss: 0.000687448657117784
step: 60, loss: 2.292828321515117e-05
step: 70, loss: 0.00029766798252239823
step: 80, loss: 3.178891711286269e-05
step: 90, loss: 2.1773246771772392e-05
step: 100, loss: 0.00017866597045212984
step: 110, loss: 0.01570861227810383
step: 120, loss: 0.0014841379597783089
step: 130, loss: 4.250325582688674e-05
step: 140, loss: 1.3060745004622731e-05
step: 150, loss: 0.0001869356638053432
step: 160, loss: 2.381129343120847e-05
step: 170, loss: 4.3028532672906294e-05
step: 180, loss: 4.711087967734784e-05
step: 190, loss: 6.301217945292592e-05
step: 200, loss: 2.5617615392548032e-05
step: 210, loss: 4.154569614911452e-05
step: 220, loss: 1.2419975973898545e-05

```

```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 230, loss: 0.061438754200935364
step: 240, loss: 3.062425457756035e-05
step: 250, loss: 4.884224108536728e-05
step: 260, loss: 3.1446626962861046e-05
step: 270, loss: 0.0003040819428861141
step: 280, loss: 0.0008586925687268376
step: 290, loss: 0.0011871502501890063
step: 300, loss: 0.0008332661236636341
step: 310, loss: 7.561594975413755e-05
step: 320, loss: 2.9797405659337528e-05
step: 330, loss: 2.989028325828258e-05
step: 340, loss: 6.252401362871751e-05
step: 350, loss: 7.0706351834815e-05
step: 360, loss: 5.1640370656969026e-05
step: 370, loss: 2.8231326723471284e-05
step: 380, loss: 1.9564769900171086e-05
step: 390, loss: 1.7925669453688897e-05
step: 400, loss: 4.2692478018580005e-05
step: 410, loss: 0.00015284327673725784
step: 420, loss: 4.420213372213766e-05
step: 430, loss: 2.0045230485266075e-05
step: 440, loss: 0.00039588805520907044
step: 450, loss: 0.0040470873937010765
step: 460, loss: 2.1956026102998294e-05
step: 470, loss: 4.8145830078283325e-05
step: 480, loss: 3.0428072932409123e-05
step: 490, loss: 0.00017083184502553195
step: 500, loss: 7.594910857733339e-05
step: 510, loss: 0.00027460232377052307
step: 520, loss: 7.411555998260155e-05
step: 530, loss: 1.948284807440359e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9369202226345084, f1=0.9345537757437071,
best_f1=0.9348623853211009
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True

```



```

loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:22, 252.93it/s]
load_f1 = 0.9380203515263645
real_f1 = 0.9377593360995851
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 252.80it/s]

```

#### 1.1.4 Amazon-Google - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Amazon-Google \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--da entry_swap \
--da span_del \
--summarize

```

```
# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt
```

```
# Step 3: run the matcher
```

```
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
 /usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
 /usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning->

```

rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
    UserWarning)
step: 0, loss: 0.8431960344314575
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06338749825954437
step: 20, loss: 0.37278902530670166
step: 30, loss: 0.3680770695209503
step: 40, loss: 0.48478832840919495
step: 50, loss: 0.29075708985328674
step: 60, loss: 0.3661310374736786
step: 70, loss: 0.21684050559997559
step: 80, loss: 0.30967697501182556
step: 90, loss: 0.3618478775024414
step: 100, loss: 0.15212391316890717
step: 110, loss: 0.25900912284851074
step: 120, loss: 0.20265163481235504
step: 130, loss: 0.2427113950252533
step: 140, loss: 0.19115811586380005
step: 150, loss: 0.2412349134683609
step: 160, loss: 0.3350849449634552
step: 170, loss: 0.09844353049993515
step: 180, loss: 0.10505782067775726
step: 190, loss: 0.33777686953544617
step: 200, loss: 0.12298814207315445
step: 210, loss: 0.4203987717628479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.623015873015873, f1=0.6294820717131474,
best_f1=0.6294820717131474
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06970542669296265
step: 10, loss: 0.1149606928229332
step: 20, loss: 0.4146394431591034
step: 30, loss: 0.06801420450210571
step: 40, loss: 0.1775522083044052
step: 50, loss: 0.2525993585586548
step: 60, loss: 0.04449371248483658
step: 70, loss: 0.11890942603349686
step: 80, loss: 0.1723499447107315
step: 90, loss: 0.07644979655742645

```

```

step: 100, loss: 0.08958374708890915
step: 110, loss: 0.07374631613492966
step: 120, loss: 0.19992172718048096
step: 130, loss: 0.23857174813747406
step: 140, loss: 0.18860016763210297
step: 150, loss: 0.14733785390853882
step: 160, loss: 0.24310347437858582
step: 170, loss: 0.24005985260009766
step: 180, loss: 0.2204592376947403
step: 190, loss: 0.19256219267845154
step: 200, loss: 0.2295251041650772
step: 210, loss: 0.19579091668128967
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6563706563706564, f1=0.6522593320235757,
best_f1=0.6522593320235757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13952043652534485
step: 10, loss: 0.1543540358543396
step: 20, loss: 0.23662731051445007
step: 30, loss: 0.058494195342063904
step: 40, loss: 0.1856282651424408
step: 50, loss: 0.1723913848400116
step: 60, loss: 0.19954903423786163
step: 70, loss: 0.08627428859472275
step: 80, loss: 0.09656919538974762
step: 90, loss: 0.12971560657024384
step: 100, loss: 0.02580191381275654
step: 110, loss: 0.1857849657535553
step: 120, loss: 0.24215666949748993
step: 130, loss: 0.26802533864974976
step: 140, loss: 0.37245240807533264
step: 150, loss: 0.22668685019016266
step: 160, loss: 0.09936882555484772
step: 170, loss: 0.19785860180854797
step: 180, loss: 0.03416601940989494
step: 190, loss: 0.10756558179855347
step: 200, loss: 0.13418817520141602
step: 210, loss: 0.14940233528614044
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6909788867562379, f1=0.6536203522504893,
best_f1=0.6536203522504893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16005071997642517
step: 10, loss: 0.029071858152747154
step: 20, loss: 0.15250687301158905
step: 30, loss: 0.035134147852659225
step: 40, loss: 0.12416243553161621
step: 50, loss: 0.05325470492243767
step: 60, loss: 0.13906042277812958
step: 70, loss: 0.15754041075706482
step: 80, loss: 0.06361810117959976
step: 90, loss: 0.12102329730987549
step: 100, loss: 0.07284817844629288
step: 110, loss: 0.11780238151550293
step: 120, loss: 0.019673019647598267
step: 130, loss: 0.1455378234386444
step: 140, loss: 0.2061183601617813
step: 150, loss: 0.0805688351392746
step: 160, loss: 0.06508427858352661
step: 170, loss: 0.032567549496889114
step: 180, loss: 0.12755803763866425
step: 190, loss: 0.07796481251716614
step: 200, loss: 0.05242907628417015
step: 210, loss: 0.01956113427877426
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7000000000000001, f1=0.6465863453815262,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14986424148082733
step: 10, loss: 0.10552242398262024
step: 20, loss: 0.04600372910499573
step: 30, loss: 0.0959228053689003

```

```
step: 40, loss: 0.2743953764438629
step: 50, loss: 0.1852804720401764
step: 60, loss: 0.05500791594386101
step: 70, loss: 0.13998067378997803
step: 80, loss: 0.027851484715938568
step: 90, loss: 0.013928717002272606
step: 100, loss: 0.03361961990594864
step: 110, loss: 0.026092583313584328
step: 120, loss: 0.020496411249041557
step: 130, loss: 0.034067731350660324
step: 140, loss: 0.09621524810791016
step: 150, loss: 0.11973302811384201
step: 160, loss: 0.09844166040420532
step: 170, loss: 0.05804756283760071
step: 180, loss: 0.2608296871185303
step: 190, loss: 0.10171359032392502
step: 200, loss: 0.14913319051265717
step: 210, loss: 0.08765518665313721
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.6802973977695167, f1=0.6865671641791046,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15842190384864807
step: 10, loss: 0.009897041134536266
step: 20, loss: 0.025266876444220543
step: 30, loss: 0.1002144068479538
step: 40, loss: 0.01013844832777977
step: 50, loss: 0.10208868980407715
step: 60, loss: 0.09502589702606201
step: 70, loss: 0.008412327617406845
step: 80, loss: 0.11941356211900711
step: 90, loss: 0.14874763786792755
step: 100, loss: 0.05903209373354912
step: 110, loss: 0.049241840839385986
step: 120, loss: 0.03976387530565262
step: 130, loss: 0.054507989436388016
step: 140, loss: 0.12711849808692932
step: 150, loss: 0.026317160576581955
step: 160, loss: 0.31366676092147827
step: 170, loss: 0.019839057698845863
step: 180, loss: 0.008148283697664738
step: 190, loss: 0.018571214750409126
step: 200, loss: 0.006332859862595797
step: 210, loss: 0.131870836019516
```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6539923954372623, f1=0.6524271844660194,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02336670272052288
step: 10, loss: 0.08701878786087036
step: 20, loss: 0.019989214837551117
step: 30, loss: 0.08717729896306992
step: 40, loss: 0.0222035963088274
step: 50, loss: 0.11919813603162766
step: 60, loss: 0.2798537015914917
step: 70, loss: 0.03160533308982849
step: 80, loss: 0.039855629205703735
step: 90, loss: 0.01717863790690899
step: 100, loss: 0.008620871230959892
step: 110, loss: 0.21733464300632477
step: 120, loss: 0.12723299860954285
step: 130, loss: 0.040358856320381165
step: 140, loss: 0.023414818570017815
step: 150, loss: 0.014260119758546352
step: 160, loss: 0.042507950216531754
step: 170, loss: 0.02392437495291233
step: 180, loss: 0.010576498694717884
step: 190, loss: 0.053177013993263245
step: 200, loss: 0.07816608995199203
step: 210, loss: 0.08086148649454117
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.6639676113360323, f1=0.6450304259634888,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08985775709152222
step: 10, loss: 0.024924518540501595
step: 20, loss: 0.01070471666753292
step: 30, loss: 0.06724116206169128
step: 40, loss: 0.07777706533670425
step: 50, loss: 0.007814999669790268
step: 60, loss: 0.1937766969203949
step: 70, loss: 0.006557994056493044
step: 80, loss: 0.029554270207881927
step: 90, loss: 0.08056368678808212

```

```

step: 100, loss: 0.012176921591162682
step: 110, loss: 0.011854193173348904
step: 120, loss: 0.03748232498764992
step: 130, loss: 0.04950571060180664
step: 140, loss: 0.01185181550681591
step: 150, loss: 0.06756898760795593
step: 160, loss: 0.07685230672359467
step: 170, loss: 0.06197286397218704
step: 180, loss: 0.04279770702123642
step: 190, loss: 0.17029327154159546
step: 200, loss: 0.2075640708208084
step: 210, loss: 0.1046818420290947
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.6563706563706564, f1=0.6540642722117203,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07132473587989807
step: 10, loss: 0.26629188656806946
step: 20, loss: 0.043065350502729416
step: 30, loss: 0.013144325464963913
step: 40, loss: 0.28125515580177307
step: 50, loss: 0.03479909524321556
step: 60, loss: 0.052955757826566696
step: 70, loss: 0.13560651242733002
step: 80, loss: 0.11277634650468826
step: 90, loss: 0.06758099049329758
step: 100, loss: 0.004634153097867966
step: 110, loss: 0.015765022486448288
step: 120, loss: 0.004062948282808065
step: 130, loss: 0.0027964820619672537
step: 140, loss: 0.013638929463922977
step: 150, loss: 0.021064113825559616
step: 160, loss: 0.017862241715192795
step: 170, loss: 0.1362294852733612
step: 180, loss: 0.06180957332253456
step: 190, loss: 0.06564126163721085
step: 200, loss: 0.029375707730650902
step: 210, loss: 0.11340184509754181
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6607538802660754, f1=0.6441441441441442,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013189823366701603
step: 10, loss: 0.03572133928537369
step: 20, loss: 0.052209850400686264
step: 30, loss: 0.011717179790139198
step: 40, loss: 0.04854502156376839
step: 50, loss: 0.016429129987955093
step: 60, loss: 0.0027973868418484926
step: 70, loss: 0.027541853487491608
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.0022449169773608446
step: 90, loss: 0.01963466964662075
step: 100, loss: 0.0014066576259210706
step: 110, loss: 0.01779305934906006
step: 120, loss: 0.04656710475683212
step: 130, loss: 0.0026457991916686296
step: 140, loss: 0.015538956969976425
step: 150, loss: 0.020767295733094215
step: 160, loss: 0.24161912500858307
step: 170, loss: 0.024698985740542412
step: 180, loss: 0.11643075197935104
step: 190, loss: 0.324695885181427
step: 200, loss: 0.09749960899353027
step: 210, loss: 0.004032778088003397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6597077244258872, f1=0.6428571428571429,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07952415198087692
step: 10, loss: 0.031069468706846237
step: 20, loss: 0.004463426303118467
step: 30, loss: 0.06303214281797409
step: 40, loss: 0.019348638132214546
step: 50, loss: 0.05018945038318634
step: 60, loss: 0.0717431828379631
step: 70, loss: 0.00797313079237938
step: 80, loss: 0.23259562253952026
step: 90, loss: 0.03217388316988945
step: 100, loss: 0.03042617067694664
step: 110, loss: 0.004736947361379862
step: 120, loss: 0.003787742927670479
step: 130, loss: 0.02124546654522419
step: 140, loss: 0.01011867169290781

```

```

step: 150, loss: 0.09974164515733719
step: 160, loss: 0.015839748084545135
step: 170, loss: 0.07382242381572723
step: 180, loss: 0.02414143830537796
step: 190, loss: 0.0066725267097353935
step: 200, loss: 0.0024306895211338997
step: 210, loss: 0.02924615889787674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6519823788546255, f1=0.6331877729257642,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001470817718654871
step: 10, loss: 0.006602882873266935
step: 20, loss: 0.004063983913511038
step: 30, loss: 0.011200565844774246
step: 40, loss: 0.01127562578767538
step: 50, loss: 0.07286252826452255
step: 60, loss: 0.003390171565115452
step: 70, loss: 0.0037258618976920843
step: 80, loss: 0.004803401418030262
step: 90, loss: 0.013861413113772869
step: 100, loss: 0.028484875336289406
step: 110, loss: 0.024447046220302582
step: 120, loss: 0.012403408996760845
step: 130, loss: 0.004020712338387966
step: 140, loss: 0.0030890083871781826
step: 150, loss: 0.007469650823622942
step: 160, loss: 0.0090648727491498
step: 170, loss: 0.006170262582600117
step: 180, loss: 0.0035844172816723585
step: 190, loss: 0.043038491159677505
step: 200, loss: 0.0811510756611824
step: 210, loss: 0.04610338807106018
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6549707602339182, f1=0.6103646833013436,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005402487702667713
step: 10, loss: 0.005178890656679869
step: 20, loss: 0.007320551201701164

```

```

step: 30, loss: 0.03793412819504738
step: 40, loss: 0.022556578740477562
step: 50, loss: 0.03913862258195877
step: 60, loss: 0.0021096467971801758
step: 70, loss: 0.1973719745874405
step: 80, loss: 0.021194541826844215
step: 90, loss: 0.06046179309487343
step: 100, loss: 0.056155674159526825
step: 110, loss: 0.002304547466337681
step: 120, loss: 0.060873258858919144
step: 130, loss: 0.016683734953403473
step: 140, loss: 0.008307205513119698
step: 150, loss: 0.012718608602881432
step: 160, loss: 0.047733910381793976
step: 170, loss: 0.11556905508041382
step: 180, loss: 0.10093779861927032
step: 190, loss: 0.0028649892192333937
step: 200, loss: 0.03693574666976929
step: 210, loss: 0.006278411019593477
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.655367231638418, f1=0.6349809885931559,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012411147356033325
step: 10, loss: 0.004558563232421875
step: 20, loss: 0.006657961290329695
step: 30, loss: 0.13689984381198883
step: 40, loss: 0.016957750543951988
step: 50, loss: 0.005513580050319433
step: 60, loss: 0.02862081676721573
step: 70, loss: 0.04367877542972565
step: 80, loss: 0.08188244700431824
step: 90, loss: 0.059610266238451004
step: 100, loss: 0.0038554170168936253
step: 110, loss: 0.08694349229335785
step: 120, loss: 0.009990314953029156
step: 130, loss: 0.0019236118532717228
step: 140, loss: 0.0008405456901527941
step: 150, loss: 0.03290148824453354
step: 160, loss: 0.017993614077568054
step: 170, loss: 0.013233796693384647
step: 180, loss: 0.012524742633104324
step: 190, loss: 0.010023406706750393
step: 200, loss: 0.0019218832021579146

```

```

step: 210, loss: 0.002078096382319927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6503759398496242, f1=0.6425855513307984,
best_f1=0.6465863453815262
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056529294699430466
step: 10, loss: 0.006122606340795755
step: 20, loss: 0.015464121475815773
step: 30, loss: 0.0011588411871343851
step: 40, loss: 0.0013890010304749012
step: 50, loss: 0.0024315989576280117
step: 60, loss: 0.004228863399475813
step: 70, loss: 0.002304516499862075
step: 80, loss: 0.002870075171813369
step: 90, loss: 0.017304005101323128
step: 100, loss: 0.01741557940840721
step: 110, loss: 0.0018398425308987498
step: 120, loss: 0.034615278244018555
step: 130, loss: 0.1335473209619522
step: 140, loss: 0.005951757542788982
step: 150, loss: 0.002238763030618429
step: 160, loss: 0.027144908905029297
step: 170, loss: 0.05994082987308502
step: 180, loss: 0.0026091388426721096
step: 190, loss: 0.008590927347540855
step: 200, loss: 0.017427794635295868
step: 210, loss: 0.030133066698908806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6510721247563354, f1=0.6367187499999999,
best_f1=0.6465863453815262
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:06, 345.93it/s]
load_f1 = 0.6924731182795699
real_f1 = 0.6768558951965066
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:16, 262.81it/s]

```

### 1.1.5 Walmart-Amazon - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Walmart-Amazon \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--da entry_swap \
--da span_del \
--summarize

```

```

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Walmart-Amazon \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8698636293411255
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.17148034274578094
step: 20, loss: 0.14790698885917664
step: 30, loss: 0.50145423412323
step: 40, loss: 0.25586703419685364
step: 50, loss: 0.3079909384250641
step: 60, loss: 0.36255815625190735
step: 70, loss: 0.1757449507713318
step: 80, loss: 0.5256521701812744
step: 90, loss: 0.23759415745735168
step: 100, loss: 0.22172382473945618
step: 110, loss: 0.23198853433132172
step: 120, loss: 0.4155820906162262
step: 130, loss: 0.36238470673561096
step: 140, loss: 0.28522562980651855
step: 150, loss: 0.2705334722995758
step: 160, loss: 0.20237945020198822
step: 170, loss: 0.3574265241622925
step: 180, loss: 0.2985917925834656
step: 190, loss: 0.11009977012872696
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6723163841807909, f1=0.6704871060171921,
best_f1=0.6704871060171921
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2146434187889099
step: 10, loss: 0.06523744016885757
step: 20, loss: 0.03407109156250954
step: 30, loss: 0.059777241200208664
step: 40, loss: 0.6061415672302246
step: 50, loss: 0.30080142617225647
step: 60, loss: 0.15401095151901245
step: 70, loss: 0.11685410141944885
step: 80, loss: 0.15729594230651855
step: 90, loss: 0.1459198147058487
step: 100, loss: 0.29454922676086426
step: 110, loss: 0.07667727023363113
step: 120, loss: 0.17570172250270844

```

```

step: 130, loss: 0.07521374523639679
step: 140, loss: 0.10061132907867432
step: 150, loss: 0.03478957712650299
step: 160, loss: 0.12827855348587036
step: 170, loss: 0.15991142392158508
step: 180, loss: 0.0846371129155159
step: 190, loss: 0.07176627963781357
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7467362924281985, f1=0.7298050139275766,
best_f1=0.7298050139275766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1008693054318428
step: 10, loss: 0.25584906339645386
step: 20, loss: 0.021796777844429016
step: 30, loss: 0.0583951473236084
step: 40, loss: 0.0367957279086113
step: 50, loss: 0.11594334989786148
step: 60, loss: 0.06080004572868347
step: 70, loss: 0.2162420004606247
step: 80, loss: 0.13782086968421936
step: 90, loss: 0.017881203442811966
step: 100, loss: 0.08432029187679291
step: 110, loss: 0.267119437456131
step: 120, loss: 0.018134871497750282
step: 130, loss: 0.011019479483366013
step: 140, loss: 0.05081452056765556
step: 150, loss: 0.1426060050725937
step: 160, loss: 0.10081348568201065
step: 170, loss: 0.04628622904419899
step: 180, loss: 0.040697988122701645
step: 190, loss: 0.057532235980033875
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7640449438202247, f1=0.7586206896551724,
best_f1=0.7586206896551724

```



```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.
```

```
    warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.03937111422419548  
step: 10, loss: 0.1316351443529129  
step: 20, loss: 0.014506548643112183  
step: 30, loss: 0.05166766047477722  
step: 40, loss: 0.0027989293448626995  
step: 50, loss: 0.09106634557247162  
step: 60, loss: 0.04711145535111427  
step: 70, loss: 0.00573381781578064  
step: 80, loss: 0.05218769609928131  
step: 90, loss: 0.09158824384212494  
step: 100, loss: 0.02332419343292713  
step: 110, loss: 0.02047480084002018  
step: 120, loss: 0.035539913922548294  
step: 130, loss: 0.15428023040294647  
step: 140, loss: 0.02811705693602562  
step: 150, loss: 0.11596475541591644  
step: 160, loss: 0.018856190145015717  
step: 170, loss: 0.02145795151591301  
step: 180, loss: 0.07545163482427597  
step: 190, loss: 0.09410702437162399
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.
```

```
    warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:  
UserWarning: Please also save or load the state of the optimizer when saving or  
loading the scheduler.
```

```
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

```
epoch 4: dev_f1=0.7688022284122563, f1=0.784741144414169,  
best_f1=0.784741144414169
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.
```

```
    warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.03263433277606964  
step: 10, loss: 0.031146129593253136  
step: 20, loss: 0.15169291198253632  
step: 30, loss: 0.015617011114954948  
step: 40, loss: 0.04085059091448784  
step: 50, loss: 0.013440804556012154  
step: 60, loss: 0.05622432753443718  
step: 70, loss: 0.005182315595448017  
step: 80, loss: 0.01929638534784317  
step: 90, loss: 0.00806386023759842  
step: 100, loss: 0.01431557908654213  
step: 110, loss: 0.009618252515792847  
step: 120, loss: 0.010016834363341331
```

```

step: 130, loss: 0.011736937798559666
step: 140, loss: 0.005855453200638294
step: 150, loss: 0.027617890387773514
step: 160, loss: 0.008124005980789661
step: 170, loss: 0.011877922341227531
step: 180, loss: 0.0347636342048645
step: 190, loss: 0.36136558651924133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7651715039577835, f1=0.7680412371134019,
best_f1=0.784741144414169
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024450572207570076
step: 10, loss: 0.000541076238732785
step: 20, loss: 0.004922105465084314
step: 30, loss: 0.003733740421012044
step: 40, loss: 0.014415755867958069
step: 50, loss: 0.023858558386564255
step: 60, loss: 0.00729749770835042
step: 70, loss: 0.07184728980064392
step: 80, loss: 0.053276851773262024
step: 90, loss: 0.1269586980342865
step: 100, loss: 0.014812707901000977
step: 110, loss: 0.2543288469314575
step: 120, loss: 0.03258489817380905
step: 130, loss: 0.007371813524514437
step: 140, loss: 0.01159267220646143
step: 150, loss: 0.14420054852962494
step: 160, loss: 0.004188696853816509
step: 170, loss: 0.12716153264045715
step: 180, loss: 0.01260356791317463
step: 190, loss: 0.022518407553434372
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7492957746478874, f1=0.7401129943502824,
best_f1=0.784741144414169
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011401467956602573
step: 10, loss: 0.030252885073423386
step: 20, loss: 0.0023623770102858543
step: 30, loss: 0.015731075778603554
step: 40, loss: 0.0538855642080307

```

```

step: 50, loss: 0.02920893207192421
step: 60, loss: 0.07240565866231918
step: 70, loss: 0.007578105665743351
step: 80, loss: 0.01766001060605049
step: 90, loss: 0.0047498405911028385
step: 100, loss: 0.003965431824326515
step: 110, loss: 0.0035069873556494713
step: 120, loss: 0.01557502243667841
step: 130, loss: 0.019048891961574554
step: 140, loss: 0.001617518486455083
step: 150, loss: 0.024683218449354172
step: 160, loss: 0.0022386652417480946
step: 170, loss: 0.0025591845624148846
step: 180, loss: 0.002502738032490015
step: 190, loss: 0.012281996197998524
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.7708894878706198, f1=0.7713498622589531,
best_f1=0.7713498622589531
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001690728240646422
step: 10, loss: 0.02053620107471943
step: 20, loss: 0.0007030722917988896
step: 30, loss: 0.0030738830100744963
step: 40, loss: 0.015189459547400475
step: 50, loss: 0.0011419416405260563
step: 60, loss: 0.001141349202953279
step: 70, loss: 0.0013277932303026319
step: 80, loss: 0.026247691363096237
step: 90, loss: 0.006884484086185694
step: 100, loss: 0.023541998118162155
step: 110, loss: 0.001068895566277206
step: 120, loss: 0.04353028163313866
step: 130, loss: 0.016906850039958954
step: 140, loss: 0.00742542976513505
step: 150, loss: 0.003584069898352027
step: 160, loss: 0.003149020718410611
step: 170, loss: 0.0010394630953669548
step: 180, loss: 0.026757746934890747
step: 190, loss: 0.10336321592330933
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7635327635327636, f1=0.7528735632183908,
best_f1=0.7713498622589531
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010276200249791145
step: 10, loss: 0.001969533506780863
step: 20, loss: 0.05063465237617493
step: 30, loss: 0.036797743290662766
step: 40, loss: 0.007242784835398197
step: 50, loss: 0.001171718817204237
step: 60, loss: 0.0017906433204188943
step: 70, loss: 0.043954331427812576
step: 80, loss: 0.0019475576700642705
step: 90, loss: 0.04004961624741554
step: 100, loss: 0.017139600589871407
step: 110, loss: 0.0017810346325859427
step: 120, loss: 0.02791823074221611
step: 130, loss: 0.002760876202955842
step: 140, loss: 0.0016722407890483737
step: 150, loss: 0.018739042803645134
step: 160, loss: 0.0004484908131416887
step: 170, loss: 0.002474126871675253
step: 180, loss: 0.008831121027469635
step: 190, loss: 0.019819894805550575
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.7731092436974789, f1=0.7586206896551724,
best_f1=0.7586206896551724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003316881484352052
step: 10, loss: 0.0014859087532386184
step: 20, loss: 0.01583317294716835
step: 30, loss: 0.008638018742203712
step: 40, loss: 0.001613516011275351
step: 50, loss: 0.0006793066277168691
step: 60, loss: 0.005173135083168745
step: 70, loss: 0.007697727996855974
step: 80, loss: 0.0015957353170961142

```

```

step: 90, loss: 0.029412809759378433
step: 100, loss: 0.01881418749690056
step: 110, loss: 0.012379329651594162
step: 120, loss: 0.055823590606451035
step: 130, loss: 0.007286197040230036
step: 140, loss: 0.0019477811874821782
step: 150, loss: 0.015512143261730671
step: 160, loss: 0.21135158836841583
step: 170, loss: 0.031190650537610054
step: 180, loss: 0.0034224919509142637
step: 190, loss: 0.005328575614839792
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.7913279132791328, f1=0.7768595041322315,
best_f1=0.7768595041322315
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004324799403548241
step: 10, loss: 0.0014934204518795013
step: 20, loss: 0.07569129765033722
step: 30, loss: 0.00416723545640707
step: 40, loss: 0.0014826871920377016
step: 50, loss: 0.0026842602528631687
step: 60, loss: 0.00126391660887748
step: 70, loss: 0.0007749879732728004
step: 80, loss: 0.0017806473188102245
step: 90, loss: 0.0060427128337323666
step: 100, loss: 0.0005441866815090179
step: 110, loss: 0.0009709172882139683
step: 120, loss: 0.0032768005039542913
step: 130, loss: 0.02135157585144043
step: 140, loss: 0.0008486948790960014
step: 150, loss: 0.008236641995608807
step: 160, loss: 0.003320635762065649
step: 170, loss: 0.00540789682418108
step: 180, loss: 0.033887457102537155
step: 190, loss: 0.0008990115602500737
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or

```

```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.8, f1=0.7606382978723404, best_f1=0.7606382978723404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001177951111458242
step: 10, loss: 0.0725637823343277
step: 20, loss: 0.013244968838989735
step: 30, loss: 0.002286963863298297
step: 40, loss: 0.004836407955735922
step: 50, loss: 0.06755346804857254
step: 60, loss: 0.0007242645951919258
step: 70, loss: 0.0009369935723952949
step: 80, loss: 0.001676345244050026
step: 90, loss: 0.004638698883354664
step: 100, loss: 0.002184379380196333
step: 110, loss: 0.0013588807778432965
step: 120, loss: 0.0037310223560780287
step: 130, loss: 0.0009007421904243529
step: 140, loss: 0.0028607568237930536
step: 150, loss: 0.0008395938784815371
step: 160, loss: 0.2349119335412979
step: 170, loss: 0.022112051025032997
step: 180, loss: 0.0013952134177088737
step: 190, loss: 0.02028024010360241
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7704918032786885, f1=0.7603305785123966,
best_f1=0.7606382978723404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11260922253131866
step: 10, loss: 0.010381744243204594
step: 20, loss: 0.0018817991949617863
step: 30, loss: 0.003127998672425747
step: 40, loss: 0.0009973536944016814
step: 50, loss: 0.0014699710300192237
step: 60, loss: 0.0010446434607729316
step: 70, loss: 0.0007674137013964355
step: 80, loss: 0.0013816851424053311
step: 90, loss: 0.0009088884107768536
step: 100, loss: 0.00206928513944149
step: 110, loss: 0.016965629532933235
step: 120, loss: 0.0035681072622537613
step: 130, loss: 0.024040699005126953

```

```

step: 140, loss: 0.0008979759877547622
step: 150, loss: 0.03461088985204697
step: 160, loss: 0.0008279440808109939
step: 170, loss: 0.0008061553235165775
step: 180, loss: 0.007382452953606844
step: 190, loss: 0.0018856576643884182
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7859078590785908, f1=0.765498652291105,
best_f1=0.7606382978723404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010163732804358006
step: 10, loss: 0.0008354974561370909
step: 20, loss: 0.006054149940609932
step: 30, loss: 0.006451786030083895
step: 40, loss: 0.0010844592470675707
step: 50, loss: 0.005293556954711676
step: 60, loss: 0.003704782808199525
step: 70, loss: 0.004452831577509642
step: 80, loss: 0.001194276032038033
step: 90, loss: 0.002280962187796831
step: 100, loss: 0.013272158801555634
step: 110, loss: 0.0020783713553100824
step: 120, loss: 0.0014861624222248793
step: 130, loss: 0.008599620312452316
step: 140, loss: 0.0008267577504739165
step: 150, loss: 0.0006086559151299298
step: 160, loss: 0.0006021769950166345
step: 170, loss: 0.0010007034288719296
step: 180, loss: 0.007059122435748577
step: 190, loss: 0.00041993887862190604
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7880434782608696, f1=0.7683923705722071,
best_f1=0.7606382978723404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009695314220152795
step: 10, loss: 0.0007294503157027066
step: 20, loss: 0.0004959626821801066
step: 30, loss: 0.0007481243228539824
step: 40, loss: 0.006126103922724724
step: 50, loss: 0.021596087142825127

```

```

step: 60, loss: 0.0008528712787665427
step: 70, loss: 0.0009740623645484447
step: 80, loss: 0.001960603054612875
step: 90, loss: 0.002887904876843095
step: 100, loss: 0.0013700019335374236
step: 110, loss: 0.000745495141018182
step: 120, loss: 0.0008046478033065796
step: 130, loss: 0.0032618555705994368
step: 140, loss: 0.007710366975516081
step: 150, loss: 0.0009585729567334056
step: 160, loss: 0.01875630021095276
step: 170, loss: 0.0009287694119848311
step: 180, loss: 0.05880947411060333
step: 190, loss: 0.012072121724486351
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7925531914893618, f1=0.7553191489361701,
best_f1=0.7606382978723404
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
2049it [00:08, 228.81it/s]
load_f1 = 0.6108786610878661
real_f1 = 0.5833333333333334
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:16, 262.11it/s]

```

## 1.2 DA TEXTUAL

### 1.2.1 Abt-Buy - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Textual/Abt-Buy \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --da entry_swap \
    --da span_shuffle \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Textual/Abt-Buy \
    --input_path input/input_small.jsonl \
    --output_path output/output_small.jsonl \
    --lm distilbert \
    --use_gpu \
    --fp16 \
    --checkpoint_path checkpoints/

```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

```

Defaults for this optimization level are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8545743823051453
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22337855398654938
step: 20, loss: 0.15953406691551208
step: 30, loss: 0.22615596652030945
step: 40, loss: 0.3200264573097229
step: 50, loss: 0.37536561489105225
step: 60, loss: 0.4386458694934845
step: 70, loss: 0.3097507655620575
step: 80, loss: 0.24418283998966217
step: 90, loss: 0.3887935280799866
step: 100, loss: 0.2276436686515808
step: 110, loss: 0.15386413037776947
step: 120, loss: 0.5031031370162964

```

```

step: 130, loss: 0.35478106141090393
step: 140, loss: 0.3351335823535919
step: 150, loss: 0.041632771492004395
step: 160, loss: 0.15934473276138306
step: 170, loss: 0.12312640249729156
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.7312348668280872, f1=0.7505938242280286,
best_f1=0.7505938242280286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10652279108762741
step: 10, loss: 0.04860396683216095
step: 20, loss: 0.1537747085094452
step: 30, loss: 0.2576073110103607
step: 40, loss: 0.12758055329322815
step: 50, loss: 0.15247842669487
step: 60, loss: 0.0973900780081749
step: 70, loss: 0.22767749428749084
step: 80, loss: 0.1866479218006134
step: 90, loss: 0.19888600707054138
step: 100, loss: 0.08635136485099792
step: 110, loss: 0.17366433143615723
step: 120, loss: 0.03592121601104736
step: 130, loss: 0.06796161085367203
step: 140, loss: 0.1646258682012558
step: 150, loss: 0.14105841517448425
step: 160, loss: 0.06630617380142212
step: 170, loss: 0.027171075344085693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7692307692307693, f1=0.7466063348416289,
best_f1=0.7466063348416289
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029002584517002106

```

```

step: 10, loss: 0.03367720916867256
step: 20, loss: 0.05842387676239014
step: 30, loss: 0.06892276555299759
step: 40, loss: 0.006174604874104261
step: 50, loss: 0.10271082818508148
step: 60, loss: 0.14647233486175537
step: 70, loss: 0.09611055254936218
step: 80, loss: 0.3483161926269531
step: 90, loss: 0.1410050243139267
step: 100, loss: 0.0317671000957489
step: 110, loss: 0.05169014260172844
step: 120, loss: 0.008622324094176292
step: 130, loss: 0.11148424446582794
step: 140, loss: 0.012012327089905739
step: 150, loss: 0.034324172884225845
step: 160, loss: 0.04693480581045151
step: 170, loss: 0.15505695343017578
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.8018648018648018, f1=0.7944572748267898,
best_f1=0.7944572748267898
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04180383309721947
step: 10, loss: 0.10803034901618958
step: 20, loss: 0.045944854617118835
step: 30, loss: 0.021658100187778473
step: 40, loss: 0.012531493790447712
step: 50, loss: 0.027757801115512848
step: 60, loss: 0.12584199011325836
step: 70, loss: 0.04044387862086296
step: 80, loss: 0.08249643445014954
step: 90, loss: 0.038909029215574265
step: 100, loss: 0.012533856555819511
step: 110, loss: 0.018362052738666534
step: 120, loss: 0.1588178277015686
step: 130, loss: 0.0697210505604744
step: 140, loss: 0.046630699187517166
step: 150, loss: 0.006120601668953896
step: 160, loss: 0.024295607581734657
step: 170, loss: 0.010350380092859268
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8126520681265207, f1=0.7703016241299303,
best_f1=0.7703016241299303
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04092922434210777
step: 10, loss: 0.03749511390924454
step: 20, loss: 0.009269134141504765
step: 30, loss: 0.017568008974194527
step: 40, loss: 0.013315224088728428
step: 50, loss: 0.01126535888761282
step: 60, loss: 0.0011219376465305686
step: 70, loss: 0.03187428042292595
step: 80, loss: 0.005233598407357931
step: 90, loss: 0.01690099574625492
step: 100, loss: 0.027023691684007645
step: 110, loss: 0.24300068616867065
step: 120, loss: 0.020465023815631866
step: 130, loss: 0.004978484474122524
step: 140, loss: 0.06196371093392372
step: 150, loss: 0.013247071765363216
step: 160, loss: 0.027179036289453506
step: 170, loss: 0.056205715984106064
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.8129675810473815, f1=0.8086124401913874,
best_f1=0.8086124401913874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0108468784019351
step: 10, loss: 0.06324928253889084
step: 20, loss: 0.004232986364513636
step: 30, loss: 0.030324026942253113
step: 40, loss: 0.0030119894072413445
step: 50, loss: 0.02789108082652092
step: 60, loss: 0.11004830151796341

```

```

step: 70, loss: 0.09605048596858978
step: 80, loss: 0.07365784794092178
step: 90, loss: 0.030866919085383415
step: 100, loss: 0.010289429686963558
step: 110, loss: 0.03685899078845978
step: 120, loss: 0.09221764653921127
step: 130, loss: 0.19775733351707458
step: 140, loss: 0.0015581067418679595
step: 150, loss: 0.21831046044826508
step: 160, loss: 0.026775743812322617
step: 170, loss: 0.03748860955238342
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.8081841432225064, f1=0.8078817733990147,
best_f1=0.8086124401913874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011969402432441711
step: 10, loss: 0.0012010191567242146
step: 20, loss: 0.007983077317476273
step: 30, loss: 0.0509440116584301
step: 40, loss: 0.023889031261205673
step: 50, loss: 0.0016187431756407022
step: 60, loss: 0.12469631433486938
step: 70, loss: 0.004623244982212782
step: 80, loss: 0.004777832422405481
step: 90, loss: 0.01662801392376423
step: 100, loss: 0.04739684239029884
step: 110, loss: 0.0004819752066396177
step: 120, loss: 0.3335941731929779
step: 130, loss: 0.11245626956224442
step: 140, loss: 0.08301965147256851
step: 150, loss: 0.071853406727314
step: 160, loss: 0.014554706402122974
step: 170, loss: 0.06234697625041008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.8229426433915212, f1=0.815165876777251,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02681545354425907
step: 10, loss: 0.005691956728696823
step: 20, loss: 0.0018088994547724724
step: 30, loss: 0.004987382795661688
step: 40, loss: 0.0038362250197678804
step: 50, loss: 0.002585840178653598
step: 60, loss: 0.016054676845669746
step: 70, loss: 0.01813267171382904
step: 80, loss: 0.0009250788716599345
step: 90, loss: 0.15381383895874023
step: 100, loss: 0.004177217371761799
step: 110, loss: 0.03926665708422661
step: 120, loss: 0.0007238058024086058
step: 130, loss: 0.0018535542767494917
step: 140, loss: 0.002892586635425687
step: 150, loss: 0.024655139073729515
step: 160, loss: 0.02495897375047207
step: 170, loss: 0.015168938785791397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.8060453400503778, f1=0.8166259168704155,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014225590275600553
step: 10, loss: 0.0034710452891886234
step: 20, loss: 0.0013833254342898726
step: 30, loss: 0.029791178181767464
step: 40, loss: 0.05497250333428383
step: 50, loss: 0.00439991382881999
step: 60, loss: 0.009638630785048008
step: 70, loss: 0.0014108758186921477
step: 80, loss: 0.03380614519119263
step: 90, loss: 0.011468474753201008
step: 100, loss: 0.03528117015957832
step: 110, loss: 0.002109994413331151
step: 120, loss: 0.0032282972242683172
step: 130, loss: 0.0011026044376194477
step: 140, loss: 0.006377214565873146
step: 150, loss: 0.0023261774331331253
step: 160, loss: 0.004381552338600159
step: 170, loss: 0.014356493018567562
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 9: dev_f1=0.7990430622009569, f1=0.7616926503340757,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010118909180164337
step: 10, loss: 0.05780376121401787
step: 20, loss: 0.0002501638955436647
step: 30, loss: 0.001403732574544847
step: 40, loss: 0.1091923713684082
step: 50, loss: 0.013299986720085144
step: 60, loss: 0.004118465352803469
step: 70, loss: 0.12607495486736298
step: 80, loss: 0.030550075694918633
step: 90, loss: 0.052603933960199356
step: 100, loss: 0.0058591170236468315
step: 110, loss: 0.045901667326688766
step: 120, loss: 0.04732998088002205
step: 130, loss: 0.01306139212101698
step: 140, loss: 0.027653060853481293
step: 150, loss: 0.0019446846563369036
step: 160, loss: 0.001332877785898745
step: 170, loss: 0.0018577477894723415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.8048192771084338, f1=0.7889908256880733,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25375282764434814
step: 10, loss: 0.00755999144166708
step: 20, loss: 0.02495800331234932
step: 30, loss: 0.0013937490293756127
step: 40, loss: 0.0024263656232506037
step: 50, loss: 0.0012721840757876635
step: 60, loss: 0.0030320349615067244
step: 70, loss: 0.0025555258616805077
step: 80, loss: 0.0005579983699135482
step: 90, loss: 0.0016141105443239212
step: 100, loss: 0.0007168562733568251
step: 110, loss: 0.0013612504117190838
step: 120, loss: 0.0012059332802891731
step: 130, loss: 0.010624425485730171
step: 140, loss: 0.010760311968624592
step: 150, loss: 0.001226237858645618
step: 160, loss: 0.0006640816573053598

```



```

step: 170, loss: 0.03331901878118515
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.8051282051282052, f1=0.8079800498753117,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012944732792675495
step: 10, loss: 0.022764824330806732
step: 20, loss: 0.022054053843021393
step: 30, loss: 0.0007337208371609449
step: 40, loss: 0.0005758744664490223
step: 50, loss: 0.002028590999543667
step: 60, loss: 0.0005892310291528702
step: 70, loss: 0.01541893184185028
step: 80, loss: 0.02958722785115242
step: 90, loss: 0.003993446007370949
step: 100, loss: 0.028948500752449036
step: 110, loss: 0.0006202982040122151
step: 120, loss: 0.07225507497787476
step: 130, loss: 0.001079193432815373
step: 140, loss: 0.0010746970074251294
step: 150, loss: 0.00412004254758358
step: 160, loss: 0.004310827236622572
step: 170, loss: 0.001112610218115151
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.8059701492537313, f1=0.8124999999999999,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005524707958102226
step: 10, loss: 0.04796228185296059
step: 20, loss: 0.0011527577880769968
step: 30, loss: 0.001422589411959052
step: 40, loss: 0.010948719456791878
step: 50, loss: 0.004789484199136496
step: 60, loss: 0.002954109339043498
step: 70, loss: 0.0010699485428631306
step: 80, loss: 0.0004215373774059117
step: 90, loss: 0.0005437016952782869
step: 100, loss: 0.00015107847866602242
step: 110, loss: 0.0019440988544374704
step: 120, loss: 0.0003781885898206383

```

```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.013558581471443176
step: 140, loss: 0.0024759515654295683
step: 150, loss: 0.0005236823344603181
step: 160, loss: 0.0004009597178082913
step: 170, loss: 0.0028988001868128777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.8117359413202935, f1=0.7990543735224587,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002540941641200334
step: 10, loss: 0.00042259509791620076
step: 20, loss: 0.01348949782550335
step: 30, loss: 0.0008045380818657577
step: 40, loss: 0.00016610162856522948
step: 50, loss: 0.12581861019134521
step: 60, loss: 0.00047754059778526425
step: 70, loss: 0.0003423937887419015
step: 80, loss: 0.0037134166341274977
step: 90, loss: 0.0004994237679056823
step: 100, loss: 0.00020399870118126273
step: 110, loss: 0.000740279327146709
step: 120, loss: 0.0003234887553844601
step: 130, loss: 0.0006329651805572212
step: 140, loss: 0.0007743805763311684
step: 150, loss: 0.0025778061244636774
step: 160, loss: 0.03782035782933235
step: 170, loss: 0.00025696007651276886
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.801980198019802, f1=0.801909307875895,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005953254294581711
step: 10, loss: 0.003382796887308359
step: 20, loss: 0.02580791525542736
step: 30, loss: 0.029580922797322273
step: 40, loss: 0.00042687045061029494
step: 50, loss: 0.0019071120768785477
step: 60, loss: 0.00012558410526253283
step: 70, loss: 0.0008642924367450178

```

```

step: 80, loss: 0.00023102283012121916
step: 90, loss: 0.00016112405864987522
step: 100, loss: 0.0005878790398128331
step: 110, loss: 0.0002805994008667767
step: 120, loss: 0.018209237605333328
step: 130, loss: 0.02924772910773754
step: 140, loss: 0.0017395588802173734
step: 150, loss: 0.008531208150088787
step: 160, loss: 0.001489709597080946
step: 170, loss: 0.0036682747304439545
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.8087167070217919, f1=0.7897196261682243,
best_f1=0.815165876777251
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:05, 338.63it/s]
load_f1 = 0.4430379746835442
real_f1 = 0.414027149321267

```

```

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:16, 262.25it/s]

```

## 1.3 DA DIRTY

### 1.3.1 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap \
  --da span_shuffle \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

```

Downloading: 100% 442/442 [00:00<00:00, 378kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 256kB/s]
Downloading: 100% 268M/268M [00:16<00:00, 16.5MB/s]
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback.

Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>", UserWarning)

step: 0, loss: 0.8200242519378662

step: 10, loss: 0.46626031398773193

step: 20, loss: 0.5836753845214844

step: 30, loss: 0.399005264043808

step: 40, loss: 0.15642154216766357

step: 50, loss: 0.10056211054325104

step: 60, loss: 0.06599494069814682

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 70, loss: 0.4321422576904297

step: 80, loss: 0.17971718311309814

step: 90, loss: 0.06367725133895874

step: 100, loss: 0.13638153672218323

step: 110, loss: 0.20653179287910461

step: 120, loss: 0.06211002543568611

step: 130, loss: 0.01748688891530037

```

step: 140, loss: 0.0042374818585813046
step: 150, loss: 0.17227381467819214
step: 160, loss: 0.12583138048648834
step: 170, loss: 0.01650121435523033
step: 180, loss: 0.008573782630264759
step: 190, loss: 0.011988564394414425
step: 200, loss: 0.007797230966389179
step: 210, loss: 0.006875847931951284
step: 220, loss: 0.005635373760014772
step: 230, loss: 0.029149631038308144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9654403567447045, f1=0.9573033707865168,
best_f1=0.9573033707865168
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010688711889088154
step: 10, loss: 0.0015156206209212542
step: 20, loss: 0.012482848018407822
step: 30, loss: 0.002241987967863679
step: 40, loss: 0.011189882643520832
step: 50, loss: 0.009492113254964352
step: 60, loss: 0.03070099838078022
step: 70, loss: 0.0335710309445858
step: 80, loss: 0.0024901784490793943
step: 90, loss: 0.009112351574003696
step: 100, loss: 0.1321423500776291
step: 110, loss: 0.11873405426740646
step: 120, loss: 0.047062948346138
step: 130, loss: 0.005567855667322874
step: 140, loss: 0.12025345116853714
step: 150, loss: 0.016256269067525864
step: 160, loss: 0.007309880573302507
step: 170, loss: 0.034665368497371674
step: 180, loss: 0.010709679685533047
step: 190, loss: 0.1993403434753418
step: 200, loss: 0.02965395152568817
step: 210, loss: 0.06689205765724182
step: 220, loss: 0.0018893342930823565
step: 230, loss: 0.019175812602043152
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.9603624009060023, f1=0.9672316384180792,
best_f1=0.9573033707865168
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07878648489713669
step: 10, loss: 0.0023372769355773926
step: 20, loss: 0.0038175873924046755
step: 30, loss: 0.07758909463882446
step: 40, loss: 0.01048717275261879
step: 50, loss: 0.010451831854879856
step: 60, loss: 0.025676941499114037
step: 70, loss: 0.0011767897522076964
step: 80, loss: 0.011190867982804775
step: 90, loss: 0.00593543378636241
step: 100, loss: 0.0009020028519444168
step: 110, loss: 0.0028774510137736797
step: 120, loss: 0.005650578998029232
step: 130, loss: 0.0008428297587670386
step: 140, loss: 0.0007226477609947324
step: 150, loss: 0.003161152359098196
step: 160, loss: 0.00600116653367877
step: 170, loss: 0.012884019874036312
step: 180, loss: 0.0013461821945384145
step: 190, loss: 0.0010568529833108187
step: 200, loss: 0.004482930060476065
step: 210, loss: 0.01430839765816927
step: 220, loss: 0.009180575609207153
step: 230, loss: 0.2587551474571228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9665178571428571, f1=0.953125, best_f1=0.953125
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018795790150761604
step: 10, loss: 0.0022762157022953033
step: 20, loss: 0.0006509179365821183
step: 30, loss: 0.0015157436719164252
step: 40, loss: 0.006061666179448366
step: 50, loss: 0.0016097198240458965
step: 60, loss: 0.0012746296124532819

```

```

step: 70, loss: 0.015750335529446602
step: 80, loss: 0.02401360124349594
step: 90, loss: 0.007675264962017536
step: 100, loss: 0.05720170959830284
step: 110, loss: 0.035807035863399506
step: 120, loss: 0.005404246971011162
step: 130, loss: 0.02978842332959175
step: 140, loss: 0.0005393003229983151
step: 150, loss: 0.0015684337122365832
step: 160, loss: 0.000978040392510593
step: 170, loss: 0.0035289719235152006
step: 180, loss: 0.2067926824092865
step: 190, loss: 0.0038890941068530083
step: 200, loss: 0.0032225316390395164
step: 210, loss: 0.03166359290480614
step: 220, loss: 0.001507320674136281
step: 230, loss: 0.0004405701474752277
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9753363228699552, f1=0.967525195968645,
best_f1=0.967525195968645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006876352708786726
step: 10, loss: 0.001319735194556415
step: 20, loss: 0.008809336461126804
step: 30, loss: 0.00030480409623123705
step: 40, loss: 0.0005578273558057845
step: 50, loss: 0.0006469156360253692
step: 60, loss: 0.0015256874030455947
step: 70, loss: 0.0004737460985779762
step: 80, loss: 0.001809444627724588
step: 90, loss: 0.005484763998538256
step: 100, loss: 0.002279409673064947
step: 110, loss: 0.0007745359325781465
step: 120, loss: 0.033504586666822433
step: 130, loss: 0.003928793128579855
step: 140, loss: 0.0006805435405112803
step: 150, loss: 0.0004572979232762009
step: 160, loss: 0.001309846295043826
step: 170, loss: 0.03353293985128403
step: 180, loss: 0.0010189013555645943

```



```

step: 190, loss: 0.0005115301464684308
step: 200, loss: 0.0004249936027918011
step: 210, loss: 0.00026906115817837417
step: 220, loss: 0.0006100229802541435
step: 230, loss: 0.0015217465115711093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9696969696969697, f1=0.967452300785634,
best_f1=0.967525195968645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005417953943833709
step: 10, loss: 0.0003747838200069964
step: 20, loss: 0.0036891703493893147
step: 30, loss: 0.001611134852282703
step: 40, loss: 0.0019658380188047886
step: 50, loss: 0.012728002853691578
step: 60, loss: 0.005314183421432972
step: 70, loss: 0.0014391435543075204
step: 80, loss: 0.0006843141745775938
step: 90, loss: 0.00035221705911681056
step: 100, loss: 0.002966204658150673
step: 110, loss: 0.02008146233856678
step: 120, loss: 0.000410585809731856
step: 130, loss: 0.017444966360926628
step: 140, loss: 0.004191531799733639
step: 150, loss: 0.00032841210486367345
step: 160, loss: 0.026084359735250473
step: 170, loss: 0.0008093299111351371
step: 180, loss: 0.0004669628688134253
step: 190, loss: 0.009427506476640701
step: 200, loss: 0.00168104306794703
step: 210, loss: 0.00032535797799937427
step: 220, loss: 0.00025113244191743433
step: 230, loss: 0.0002341564977541566
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9724366041896362, f1=0.9711751662971175,
best_f1=0.967525195968645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003097973996773362
step: 10, loss: 0.000221114547457546
step: 20, loss: 0.0004647238238248974

```

```

step: 30, loss: 0.003243308514356613
step: 40, loss: 0.0002839773369487375
step: 50, loss: 0.0002384917315794155
step: 60, loss: 0.014914687722921371
step: 70, loss: 0.001564245205372572
step: 80, loss: 0.00034138085902668536
step: 90, loss: 0.0008821145165711641
step: 100, loss: 0.00020999864500481635
step: 110, loss: 0.002149408683180809
step: 120, loss: 0.016917772591114044
step: 130, loss: 0.0009585028165020049
step: 140, loss: 0.000171675332239829
step: 150, loss: 0.00851071160286665
step: 160, loss: 0.0010103489039465785
step: 170, loss: 0.00023441347002517432
step: 180, loss: 0.0003636385954450816
step: 190, loss: 0.0002746787795331329
step: 200, loss: 0.000287833740003407
step: 210, loss: 0.00019193576008547097
step: 220, loss: 0.009821443818509579
step: 230, loss: 0.015563332475721836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9764837625979844, f1=0.967305524239008,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017951952293515205
step: 10, loss: 0.005260947160422802
step: 20, loss: 0.0013980548828840256
step: 30, loss: 0.00044494884787127376
step: 40, loss: 0.000596154248341918
step: 50, loss: 0.0020534577779471874
step: 60, loss: 0.0004949082504026592
step: 70, loss: 0.0006298245280049741
step: 80, loss: 0.00038459236384369433
step: 90, loss: 0.00027841003611683846
step: 100, loss: 0.0002497523673810065
step: 110, loss: 0.013347573578357697
step: 120, loss: 0.00011819331848528236
step: 130, loss: 0.0010345238260924816
step: 140, loss: 0.0005453634657897055

```

```

step: 150, loss: 0.00011698593880282715
step: 160, loss: 0.0013099396601319313
step: 170, loss: 0.00034965696977451444
step: 180, loss: 0.0019152287859469652
step: 190, loss: 0.0011393807362765074
step: 200, loss: 0.0012294221669435501
step: 210, loss: 0.00018645625095814466
step: 220, loss: 0.0001969263976207003
step: 230, loss: 0.0011690708342939615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.974472807991121, f1=0.9675977653631285,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016865757061168551
step: 10, loss: 0.00038476160261780024
step: 20, loss: 0.0015726074343547225
step: 30, loss: 0.0004050248535349965
step: 40, loss: 0.00019598555809352547
step: 50, loss: 0.002486046403646469
step: 60, loss: 0.00011861210077768192
step: 70, loss: 0.0004924738896079361
step: 80, loss: 0.01604955457150936
step: 90, loss: 0.004786915145814419
step: 100, loss: 0.0034825773909687996
step: 110, loss: 0.00010734663374023512
step: 120, loss: 0.04162903130054474
step: 130, loss: 0.00014339873450808227
step: 140, loss: 0.01129674632102251
step: 150, loss: 8.610907389083877e-05
step: 160, loss: 0.0028588254936039448
step: 170, loss: 0.0003308104060124606
step: 180, loss: 0.00045929441694170237
step: 190, loss: 0.00015189440455287695
step: 200, loss: 0.00015808202442713082
step: 210, loss: 0.0001419859763700515
step: 220, loss: 0.030746495351195335
step: 230, loss: 0.00012195489398436621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9744160177975528, f1=0.9656699889258028,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010979612125083804
step: 10, loss: 0.00010556067718425766
step: 20, loss: 0.00011619454744504765
step: 30, loss: 0.00040943664498627186
step: 40, loss: 0.00011075264046667144
step: 50, loss: 0.00036548959906212986
step: 60, loss: 0.032787010073661804
step: 70, loss: 0.00016079885244835168
step: 80, loss: 0.00010519666830077767
step: 90, loss: 0.00017205654876306653
step: 100, loss: 0.0008895557257346809
step: 110, loss: 0.00012211281864438206
step: 120, loss: 0.024701477959752083
step: 130, loss: 0.002363368170335889
step: 140, loss: 0.014034409075975418
step: 150, loss: 0.0003512392286211252
step: 160, loss: 0.0006816584500484169
step: 170, loss: 0.00021366823057178408
step: 180, loss: 0.0008002867689356208
step: 190, loss: 0.00014224743063095957
step: 200, loss: 0.0005348403356038034
step: 210, loss: 0.00010079279309138656
step: 220, loss: 0.0005025114514864981
step: 230, loss: 0.00017391466826666147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9730941704035874, f1=0.9664429530201343,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015233979502227157
step: 10, loss: 0.00010909230331890285
step: 20, loss: 6.636550097027794e-05
step: 30, loss: 0.0002812446909956634
step: 40, loss: 0.006061608903110027
step: 50, loss: 0.0023192374501377344
step: 60, loss: 0.00048649206291884184
step: 70, loss: 0.00012332941696513444
step: 80, loss: 0.0001454036682844162
step: 90, loss: 7.453739817719907e-05
step: 100, loss: 0.00014635267143603414
step: 110, loss: 0.0011107209138572216
step: 120, loss: 0.00010277688852511346
step: 130, loss: 4.731634908239357e-05
step: 140, loss: 5.660793249262497e-05

```

```

step: 150, loss: 5.37783962499816e-05
step: 160, loss: 0.00014059527893550694
step: 170, loss: 0.0038017970509827137
step: 180, loss: 0.00012660800712183118
step: 190, loss: 8.62323577166535e-05
step: 200, loss: 7.268234912771732e-05
step: 210, loss: 6.548529927385971e-05
step: 220, loss: 6.604234658880159e-05
step: 230, loss: 8.142289152601734e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9730941704035874, f1=0.9641255605381166,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010613713675411418
step: 10, loss: 0.00018908428319264203
step: 20, loss: 4.341257590567693e-05
step: 30, loss: 0.0011747335083782673
step: 40, loss: 5.486582813318819e-05
step: 50, loss: 5.867229265277274e-05
step: 60, loss: 0.016517378389835358
step: 70, loss: 4.872736462857574e-05
step: 80, loss: 3.803307481575757e-05
step: 90, loss: 6.425270112231374e-05
step: 100, loss: 4.2402694816701114e-05
step: 110, loss: 0.028938954696059227
step: 120, loss: 6.328327435767278e-05
step: 130, loss: 0.00020365281670819968
step: 140, loss: 6.428374035749584e-05
step: 150, loss: 3.5801102058030665e-05
step: 160, loss: 6.159661279525608e-05
step: 170, loss: 6.193958688527346e-05
step: 180, loss: 6.904332985868677e-05
step: 190, loss: 5.975205931463279e-05
step: 200, loss: 0.006688095163553953
step: 210, loss: 5.009834421798587e-05
step: 220, loss: 0.0008789275889284909
step: 230, loss: 0.0001327779609709978
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9762174405436014, f1=0.9684684684684683,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003051165258511901
step: 10, loss: 0.0001906997204059735
step: 20, loss: 6.562346970895305e-05
step: 30, loss: 0.0068746632896363735
step: 40, loss: 0.00044943674583919346
step: 50, loss: 0.00010726973414421082
step: 60, loss: 7.635764632141218e-05
step: 70, loss: 0.008845350705087185
step: 80, loss: 4.427518433658406e-05
step: 90, loss: 4.1019437048817053e-05
step: 100, loss: 5.315287853591144e-05
step: 110, loss: 6.253543688217178e-05
step: 120, loss: 8.00255875219591e-05
step: 130, loss: 0.0001236992102349177
step: 140, loss: 0.012717628851532936
step: 150, loss: 0.0003930531383957714
step: 160, loss: 7.73088977439329e-05
step: 170, loss: 9.171001147478819e-05
step: 180, loss: 0.00015242869267240167
step: 190, loss: 4.607137816492468e-05
step: 200, loss: 0.0003357390232849866
step: 210, loss: 0.00010453250433783978
step: 220, loss: 0.0002075391821563244
step: 230, loss: 5.344301462173462e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9754464285714286, f1=0.9642058165548099,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009647135157138109
step: 10, loss: 3.083324554609135e-05
step: 20, loss: 0.00043498858576640487
step: 30, loss: 9.272080933442339e-05
step: 40, loss: 5.152609446668066e-05
step: 50, loss: 0.007858074270188808
step: 60, loss: 3.711304452735931e-05
step: 70, loss: 5.368432539398782e-05
step: 80, loss: 0.00017821192159317434
step: 90, loss: 0.0004448599647730589
step: 100, loss: 0.004931560717523098
step: 110, loss: 0.00011713044659700245
step: 120, loss: 5.2689127187477425e-05
step: 130, loss: 8.467159204883501e-05
step: 140, loss: 4.2242114432156086e-05

```

```

step: 150, loss: 7.076218753354624e-05
step: 160, loss: 3.789090260397643e-05
step: 170, loss: 0.0006595429149456322
step: 180, loss: 5.894050627830438e-05
step: 190, loss: 0.00032984826248139143
step: 200, loss: 3.773135904339142e-05
step: 210, loss: 7.561739766970277e-05
step: 220, loss: 0.00015136321599129587
step: 230, loss: 2.4687016775715165e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9763779527559054, f1=0.9662921348314607,
best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.86106148653198e-05
step: 10, loss: 9.106894867727533e-05
step: 20, loss: 7.9500547144562e-05
step: 30, loss: 6.266808486543596e-05
step: 40, loss: 8.754779992159456e-05
step: 50, loss: 9.975602006306872e-05
step: 60, loss: 3.237877899664454e-05
step: 70, loss: 6.848712655482814e-05
step: 80, loss: 4.272722071618773e-05
step: 90, loss: 2.8084068617317826e-05
step: 100, loss: 5.1018821977777407e-05
step: 110, loss: 4.9308317102259025e-05
step: 120, loss: 7.205225119832903e-05
step: 130, loss: 7.932162407087162e-05
step: 140, loss: 0.0025108580011874437
step: 150, loss: 0.00017610102077014744
step: 160, loss: 0.000275413622148335
step: 170, loss: 3.518791709211655e-05
step: 180, loss: 4.9840549763757735e-05
step: 190, loss: 7.711940997978672e-05
step: 200, loss: 4.565965718938969e-05
step: 210, loss: 0.003211226314306259
step: 220, loss: 0.0010746577754616737
step: 230, loss: 7.154534250730649e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9752808988764046, f1=0.9685393258426966,
best_f1=0.967305524239008
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32

```

master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

2473it [00:10, 227.60it/s]

load\_f1 = 0.976324689966178

real\_f1 = 0.9717514124293786

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

4398it [00:18, 234.34it/s]

### 1.3.2 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Dirty/DBLP-GoogleScholar \
        --batch_size 32 \
        --max_len 256 \
```



```

--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--da span_shuffle \
--da attr_del \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback.

Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
```

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.7924340963363647
step: 10, loss: 0.3947617709636688
step: 20, loss: 0.4884870946407318
step: 30, loss: 0.42896145582199097
step: 40, loss: 0.3188297152519226
step: 50, loss: 0.17287829518318176
step: 60, loss: 0.17884114384651184
step: 70, loss: 0.05415474995970726
step: 80, loss: 0.14824658632278442
step: 90, loss: 0.1214735358953476
step: 100, loss: 0.37167200446128845
step: 110, loss: 0.1276944875717163
step: 120, loss: 0.09182054549455643
step: 130, loss: 0.03831883519887924
step: 140, loss: 0.15987806022167206
step: 150, loss: 0.03041173331439495
step: 160, loss: 0.15955951809883118
step: 170, loss: 0.18028731644153595
step: 180, loss: 0.11246433109045029
step: 190, loss: 0.05629245936870575
step: 200, loss: 0.14933402836322784
step: 210, loss: 0.10614348202943802
step: 220, loss: 0.09305387735366821
```

```
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 230, loss: 0.12745948135852814
step: 240, loss: 0.17922845482826233
step: 250, loss: 0.08087079226970673
step: 260, loss: 0.03144403174519539
step: 270, loss: 0.024731535464525223
step: 280, loss: 0.14364022016525269
step: 290, loss: 0.06302449852228165
step: 300, loss: 0.15584665536880493
step: 310, loss: 0.059777624905109406
step: 320, loss: 0.06998135894536972
step: 330, loss: 0.16857632994651794
step: 340, loss: 0.09677182137966156
```

```

step: 350, loss: 0.17878122627735138
step: 360, loss: 0.07559756934642792
step: 370, loss: 0.08036951720714569
step: 380, loss: 0.17642970383167267
step: 390, loss: 0.02285381779074669
step: 400, loss: 0.04089657962322235
step: 410, loss: 0.09269697219133377
step: 420, loss: 0.014485053718090057
step: 430, loss: 0.030164165422320366
step: 440, loss: 0.1296166479587555
step: 450, loss: 0.012352201156318188
step: 460, loss: 0.02461559697985649
step: 470, loss: 0.18937885761260986
step: 480, loss: 0.1806405782699585
step: 490, loss: 0.019425390288233757
step: 500, loss: 0.026326531544327736
step: 510, loss: 0.12108296900987625
step: 520, loss: 0.11681029200553894
step: 530, loss: 0.09328078478574753
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9297445255474452, f1=0.932118451025057,
best_f1=0.932118451025057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12137282639741898
step: 10, loss: 0.153101846575737
step: 20, loss: 0.19629491865634918
step: 30, loss: 0.05961792171001434
step: 40, loss: 0.011855051852762699
step: 50, loss: 0.11270413547754288
step: 60, loss: 0.25558018684387207
step: 70, loss: 0.15125778317451477
step: 80, loss: 0.01196263451129198
step: 90, loss: 0.027181409299373627
step: 100, loss: 0.17139798402786255
step: 110, loss: 0.019770348444581032
step: 120, loss: 0.056498173624277115
step: 130, loss: 0.015419475734233856
step: 140, loss: 0.05576916038990021
step: 150, loss: 0.06208614632487297
step: 160, loss: 0.051772382110357285

```

```

step: 170, loss: 0.08119552582502365
step: 180, loss: 0.016709474846720695
step: 190, loss: 0.03289976716041565
step: 200, loss: 0.0225289948284626
step: 210, loss: 0.034634850919246674
step: 220, loss: 0.1113491952419281
step: 230, loss: 0.07625774294137955
step: 240, loss: 0.07376039028167725
step: 250, loss: 0.032654400914907455
step: 260, loss: 0.11509104818105698
step: 270, loss: 0.08509211242198944
step: 280, loss: 0.18906868994235992
step: 290, loss: 0.23924672603607178
step: 300, loss: 0.035842880606651306
step: 310, loss: 0.06773463636636734
step: 320, loss: 0.23834499716758728
step: 330, loss: 0.018761154264211655
step: 340, loss: 0.06468501687049866
step: 350, loss: 0.051904208958148956
step: 360, loss: 0.10467270016670227
step: 370, loss: 0.011756528168916702
step: 380, loss: 0.06751126050949097
step: 390, loss: 0.03651875630021095
step: 400, loss: 0.06341345608234406
step: 410, loss: 0.000588324386626482
step: 420, loss: 0.06079475209116936
step: 430, loss: 0.013349190354347229
step: 440, loss: 0.009152678772807121
step: 450, loss: 0.013310877606272697
step: 460, loss: 0.2139609456062317
step: 470, loss: 0.014027138240635395
step: 480, loss: 0.2087600976228714
step: 490, loss: 0.012762458063662052
step: 500, loss: 0.012025384232401848
step: 510, loss: 0.1439608633518219
step: 520, loss: 0.07227028906345367
step: 530, loss: 0.20393632352352142
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9386248269497001, f1=0.9390187987161852,
best_f1=0.9390187987161852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01606808602809906
step: 10, loss: 0.05685438960790634
step: 20, loss: 0.2500026524066925
step: 30, loss: 0.13682445883750916
step: 40, loss: 0.010674075223505497
step: 50, loss: 0.048090629279613495
step: 60, loss: 0.00870904978364706
step: 70, loss: 0.07452388107776642
step: 80, loss: 0.00592036684975028
step: 90, loss: 0.04233719781041145
step: 100, loss: 0.00979859009385109
step: 110, loss: 0.06721377372741699
step: 120, loss: 0.031352508813142776
step: 130, loss: 0.03783497214317322
step: 140, loss: 0.013266276568174362
step: 150, loss: 0.011696407571434975
step: 160, loss: 0.005533034447580576
step: 170, loss: 0.00747678941115737
step: 180, loss: 0.03643205016851425
step: 190, loss: 0.02522759884595871
step: 200, loss: 0.033964235335588455
step: 210, loss: 0.07605462521314621
step: 220, loss: 0.005254677496850491
step: 230, loss: 0.022935327142477036
step: 240, loss: 0.0034046343062072992
step: 250, loss: 0.010891018435359001
step: 260, loss: 0.05766819417476654
step: 270, loss: 0.005244705826044083
step: 280, loss: 0.044940125197172165
step: 290, loss: 0.10504364222288132
step: 300, loss: 0.14007605612277985
step: 310, loss: 0.17446747422218323
step: 320, loss: 0.16685298085212708
step: 330, loss: 0.004480480682104826
step: 340, loss: 0.003184025641530752
step: 350, loss: 0.04275966435670853
step: 360, loss: 0.004597559105604887
step: 370, loss: 0.007142967544496059
step: 380, loss: 0.05178771913051605
step: 390, loss: 0.008448095992207527
step: 400, loss: 0.009299002587795258
step: 410, loss: 0.029846275225281715
step: 420, loss: 0.05296264961361885
step: 430, loss: 0.05405338108539581
step: 440, loss: 0.014131134375929832
step: 450, loss: 0.020755309611558914
step: 460, loss: 0.05290692672133446
```

```

step: 470, loss: 0.0019319010898470879
step: 480, loss: 0.012719456106424332
step: 490, loss: 0.03529328480362892
step: 500, loss: 0.10598147660493851
step: 510, loss: 0.011860677972435951
step: 520, loss: 0.0014701483305543661
step: 530, loss: 0.03225591778755188
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9373549883990719, f1=0.9324699352451433,
best_f1=0.9390187987161852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012171355076134205
step: 10, loss: 0.004736543167382479
step: 20, loss: 0.14992666244506836
step: 30, loss: 0.005649361293762922
step: 40, loss: 0.00836693774908781
step: 50, loss: 0.03630661964416504
step: 60, loss: 0.001387074589729309
step: 70, loss: 0.0011543792206794024
step: 80, loss: 0.029037902131676674
step: 90, loss: 0.015229837968945503
step: 100, loss: 0.0020811285357922316
step: 110, loss: 0.0156936664134264
step: 120, loss: 0.13266445696353912
step: 130, loss: 0.0018772121984511614
step: 140, loss: 0.0740610733628273
step: 150, loss: 0.010952884331345558
step: 160, loss: 0.03927246853709221
step: 170, loss: 0.01329164206981659
step: 180, loss: 0.024860627949237823
step: 190, loss: 0.01275709830224514
step: 200, loss: 0.019674882292747498
step: 210, loss: 0.07067697495222092
step: 220, loss: 0.01411417219787836
step: 230, loss: 0.2325206995010376
step: 240, loss: 0.060051899403333664
step: 250, loss: 0.0028199071530252695
step: 260, loss: 0.144662007689476
step: 270, loss: 0.061236754059791565
step: 280, loss: 0.009460153989493847
step: 290, loss: 0.0714663565158844
step: 300, loss: 0.01910777948796749
step: 310, loss: 0.008705325424671173
step: 320, loss: 0.062328748404979706

```

```

step: 330, loss: 0.09204603731632233
step: 340, loss: 0.037929341197013855
step: 350, loss: 0.0021259640343487263
step: 360, loss: 0.013971536420285702
step: 370, loss: 0.09315922111272812
step: 380, loss: 0.024808108806610107
step: 390, loss: 0.08191566169261932
step: 400, loss: 0.0158956591039896
step: 410, loss: 0.10930648446083069
step: 420, loss: 0.01000642403960228
step: 430, loss: 0.027875637635588646
step: 440, loss: 0.06971518695354462
step: 450, loss: 0.00697863707318902
step: 460, loss: 0.0010637290542945266
step: 470, loss: 0.003108866047114134
step: 480, loss: 0.003673589089885354
step: 490, loss: 0.03477337211370468
step: 500, loss: 0.013711758889257908
step: 510, loss: 0.020165178924798965
step: 520, loss: 0.1096460148692131
step: 530, loss: 0.001162372762337327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9347826086956521, f1=0.9335205992509362,
best_f1=0.9390187987161852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027229958213865757
step: 10, loss: 0.0068163685500621796
step: 20, loss: 0.0023310875985771418
step: 30, loss: 0.0019155541667714715
step: 40, loss: 0.027472209185361862
step: 50, loss: 0.04501219093799591
step: 60, loss: 0.0019864696078002453
step: 70, loss: 0.0015418855473399162
step: 80, loss: 0.0028381722513586283
step: 90, loss: 0.028672611340880394
step: 100, loss: 0.0014638743596151471
step: 110, loss: 0.05338190123438835
step: 120, loss: 0.015156934969127178
step: 130, loss: 0.0030899320263415575
step: 140, loss: 0.020191771909594536
step: 150, loss: 0.00044782552868127823
step: 160, loss: 0.06828733533620834
step: 170, loss: 0.08434086292982101
step: 180, loss: 0.0027950406074523926

```

```

step: 190, loss: 0.0005304748192429543
step: 200, loss: 0.006171738728880882
step: 210, loss: 0.00500827468931675
step: 220, loss: 0.00035116347135044634
step: 230, loss: 0.0014683665940538049
step: 240, loss: 0.0051633683033287525
step: 250, loss: 0.010387495160102844
step: 260, loss: 0.0071184406988322735
step: 270, loss: 0.03539174050092697
step: 280, loss: 0.05880064144730568
step: 290, loss: 0.0011906895088031888
step: 300, loss: 0.11656875908374786
step: 310, loss: 0.002945639891549945
step: 320, loss: 0.06814989447593689
step: 330, loss: 0.0035616285167634487
step: 340, loss: 0.00747778220102191
step: 350, loss: 0.016095558181405067
step: 360, loss: 0.0033046752214431763
step: 370, loss: 0.001705483184196055
step: 380, loss: 0.19202232360839844
step: 390, loss: 0.009772997349500656
step: 400, loss: 0.044144388288259506
step: 410, loss: 0.0019212262704968452
step: 420, loss: 0.00439325999468565
step: 430, loss: 0.039848241955041885
step: 440, loss: 0.10012124478816986
step: 450, loss: 0.0008200456504710019
step: 460, loss: 0.04011707380414009
step: 470, loss: 0.008224722929298878
step: 480, loss: 0.032340265810489655
step: 490, loss: 0.015939179807901382
step: 500, loss: 0.008958726190030575
step: 510, loss: 0.024805180728435516
step: 520, loss: 0.08609679341316223
step: 530, loss: 0.010546667501330376
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.937962962962963, f1=0.934752429430819,
best_f1=0.9390187987161852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009689579019322991
step: 10, loss: 0.004261981695890427
step: 20, loss: 0.011749347671866417
step: 30, loss: 0.0036113846581429243
step: 40, loss: 0.12909996509552002

```



step: 50, loss: 0.0008722793427295983  
step: 60, loss: 0.0006015025428496301  
step: 70, loss: 0.024879029020667076  
step: 80, loss: 0.002898812759667635  
step: 90, loss: 0.016283219680190086  
step: 100, loss: 0.05387892201542854  
step: 110, loss: 0.02748638391494751  
step: 120, loss: 0.002806937787681818  
step: 130, loss: 0.000858386920299381  
step: 140, loss: 0.004144562408328056  
step: 150, loss: 0.0016743142623454332  
step: 160, loss: 8.146886102622375e-05  
step: 170, loss: 0.00017568186740390956  
step: 180, loss: 0.011711717583239079  
step: 190, loss: 0.018150798976421356  
step: 200, loss: 0.002386155305430293  
step: 210, loss: 0.021018704399466515  
step: 220, loss: 0.00038170034531503916  
step: 230, loss: 0.0009191519347950816  
step: 240, loss: 0.0005393904866650701  
step: 250, loss: 0.011950196698307991  
step: 260, loss: 0.0039123790338635445  
step: 270, loss: 0.0004893392906524241  
step: 280, loss: 0.009136153385043144  
step: 290, loss: 0.0005062369164079428  
step: 300, loss: 0.002067530993372202  
step: 310, loss: 0.0014404437970370054  
step: 320, loss: 0.003739834064617753  
step: 330, loss: 0.013454077765345573  
step: 340, loss: 0.01779470033943653  
step: 350, loss: 0.11997680366039276  
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0  
step: 360, loss: 0.00010230653424514458  
step: 370, loss: 0.04765702411532402  
step: 380, loss: 0.09742055088281631  
step: 390, loss: 0.01835090108215809  
step: 400, loss: 0.0004502933588810265  
step: 410, loss: 0.0007237779209390283  
step: 420, loss: 0.0048798611387610435  
step: 430, loss: 0.00622681574895978  
step: 440, loss: 0.08324714004993439  
step: 450, loss: 0.005255580879747868  
step: 460, loss: 0.008102252148091793  
step: 470, loss: 0.051399946212768555  
step: 480, loss: 0.01806020177900791  
step: 490, loss: 0.00017909800226334482  
step: 500, loss: 0.003324448596686125  
step: 510, loss: 0.0003692099417094141

```

step: 520, loss: 0.007126143667846918
step: 530, loss: 0.0051576863043010235
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9398601398601399, f1=0.9323447636700649,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004868716932833195
step: 10, loss: 0.03890702873468399
step: 20, loss: 0.005268689710646868
step: 30, loss: 0.029959792271256447
step: 40, loss: 0.0003766703011933714
step: 50, loss: 0.039674073457717896
step: 60, loss: 0.10129247605800629
step: 70, loss: 0.0011409843573346734
step: 80, loss: 0.0011874716728925705
step: 90, loss: 0.007007824722677469
step: 100, loss: 0.0014165412867441773
step: 110, loss: 0.00036818336229771376
step: 120, loss: 0.0036374316550791264
step: 130, loss: 0.0004596910148393363
step: 140, loss: 0.0030796872451901436
step: 150, loss: 0.1467626690864563
step: 160, loss: 0.0011430061422288418
step: 170, loss: 0.000665091909468174
step: 180, loss: 0.009063072502613068
step: 190, loss: 0.00020677967404481024
step: 200, loss: 0.13560597598552704
step: 210, loss: 0.018231209367513657
step: 220, loss: 0.0003376907261554152
step: 230, loss: 0.0001805917709134519
step: 240, loss: 0.001783400890417397
step: 250, loss: 0.006918884813785553
step: 260, loss: 0.02267363667488098
step: 270, loss: 0.0004730421642307192
step: 280, loss: 0.05160146206617355
step: 290, loss: 0.056359097361564636
step: 300, loss: 0.0020194542594254017
step: 310, loss: 0.0002080803969874978
step: 320, loss: 0.05436180531978607
step: 330, loss: 0.02567763812839985

```

```
step: 340, loss: 0.02326939068734646
step: 350, loss: 0.008485471829771996
step: 360, loss: 0.0035190305206924677
step: 370, loss: 0.03249429911375046
step: 380, loss: 0.0008754876325838268
step: 390, loss: 0.0001216260643559508
step: 400, loss: 0.00013189656601753086
step: 410, loss: 0.006189845036715269
step: 420, loss: 0.005963234696537256
step: 430, loss: 6.992222915869206e-05
step: 440, loss: 0.0003955377615056932
step: 450, loss: 0.03051000088453293
step: 460, loss: 0.003909096121788025
step: 470, loss: 0.03886113315820694
step: 480, loss: 0.001069714897312224
step: 490, loss: 0.00393102178350091
step: 500, loss: 0.003533731447532773
step: 510, loss: 0.0033190157264471054
step: 520, loss: 0.0023180567659437656
step: 530, loss: 0.0005618015420623124
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9397031539888682, f1=0.9321723189734189,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029650921002030373
step: 10, loss: 0.02076839469373226
step: 20, loss: 0.012539948336780071
step: 30, loss: 0.0022554798051714897
step: 40, loss: 0.0011312610004097223
step: 50, loss: 0.00626341812312603
step: 60, loss: 0.00027354154735803604
step: 70, loss: 0.013162250630557537
step: 80, loss: 0.00016882334602996707
step: 90, loss: 0.00014243512123357505
step: 100, loss: 0.000653305382002145
step: 110, loss: 0.0008760016062296927
step: 120, loss: 0.00014555959205608815
step: 130, loss: 0.0059199552051723
step: 140, loss: 4.5215794671094045e-05
step: 150, loss: 0.0006819535628892481
step: 160, loss: 0.00036662950878962874
step: 170, loss: 0.005104451440274715
step: 180, loss: 0.03616327792406082
step: 190, loss: 0.0034618228673934937
```

```

step: 200, loss: 0.000638489902485162
step: 210, loss: 0.004759117960929871
step: 220, loss: 0.00036474669468589127
step: 230, loss: 0.0002710307890083641
step: 240, loss: 0.010213850066065788
step: 250, loss: 0.0006564627401530743
step: 260, loss: 0.23004524409770966
step: 270, loss: 0.00029790657572448254
step: 280, loss: 0.0030952095985412598
step: 290, loss: 0.018865084275603294
step: 300, loss: 0.0007556467899121344
step: 310, loss: 0.07915867865085602
step: 320, loss: 0.0007243255968205631
step: 330, loss: 0.024342797696590424
step: 340, loss: 0.0027683647349476814
step: 350, loss: 0.011704615317285061
step: 360, loss: 0.024518514052033424
step: 370, loss: 0.0012785529252141714
step: 380, loss: 0.00723070977255702
step: 390, loss: 0.003281814744696021
step: 400, loss: 0.05248570442199707
step: 410, loss: 0.004022295121103525
step: 420, loss: 0.0024714746978133917
step: 430, loss: 0.012247320264577866
step: 440, loss: 0.007677037734538317
step: 450, loss: 0.0025114656891673803
step: 460, loss: 0.00028160354122519493
step: 470, loss: 0.02422620728611946
step: 480, loss: 0.0004415986477397382
step: 490, loss: 0.000787450815550983
step: 500, loss: 0.048019785434007645
step: 510, loss: 0.00045690874685533345
step: 520, loss: 0.00337756285443902
step: 530, loss: 0.017505241557955742
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9376163873370577, f1=0.9325946445060019,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003375774249434471
step: 10, loss: 0.0007984685362316668
step: 20, loss: 0.0008796451147645712
step: 30, loss: 9.17953802854754e-05
step: 40, loss: 0.0382673516869545
step: 50, loss: 0.00039987594936974347

```

step: 60, loss: 0.0027541264425963163  
step: 70, loss: 0.12773258984088898  
step: 80, loss: 7.829633250366896e-05  
step: 90, loss: 0.01585717871785164  
step: 100, loss: 0.0014439948135986924  
step: 110, loss: 0.0010764404432848096  
step: 120, loss: 6.592022691620514e-05  
step: 130, loss: 0.0005102024879306555  
step: 140, loss: 0.014213272370398045  
step: 150, loss: 0.0002763134252745658  
step: 160, loss: 0.00025513715809211135  
step: 170, loss: 0.0001600733958184719  
step: 180, loss: 0.0006177153554745018  
step: 190, loss: 0.0015965915517881513  
step: 200, loss: 0.000923736544791609  
step: 210, loss: 5.862920079380274e-05  
step: 220, loss: 0.0001997347571887076  
step: 230, loss: 8.147265180014074e-05  
step: 240, loss: 0.0002190856175730005  
step: 250, loss: 5.070922270533629e-05  
step: 260, loss: 0.0937836691737175  
step: 270, loss: 0.010743875987827778  
step: 280, loss: 6.69958972139284e-05  
step: 290, loss: 4.0208902646554634e-05  
step: 300, loss: 0.00024042643781285733  
step: 310, loss: 0.00016732324729673564  
step: 320, loss: 0.0007805753266438842  
step: 330, loss: 9.312847396358848e-05  
step: 340, loss: 0.00022500805789604783  
step: 350, loss: 0.00011662863107630983  
step: 360, loss: 0.0024206729140132666  
step: 370, loss: 0.0001309320068685338  
step: 380, loss: 5.024936763220467e-05  
step: 390, loss: 5.370291546569206e-05  
step: 400, loss: 0.042549196630716324  
step: 410, loss: 0.02157129906117916  
step: 420, loss: 0.0003784717991948128  
step: 430, loss: 2.386388041486498e-05  
step: 440, loss: 0.03143967315554619  
step: 450, loss: 7.25126956240274e-05  
step: 460, loss: 0.00411549536511302  
step: 470, loss: 0.0002764117962215096  
step: 480, loss: 8.903525304049253e-05  
step: 490, loss: 0.00021612962882500142  
step: 500, loss: 0.008023005910217762  
step: 510, loss: 0.0009444364113733172  
step: 520, loss: 0.005291655659675598  
step: 530, loss: 0.00839049182832241

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9329608938547487, f1=0.9307298930729893,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005063308635726571
step: 10, loss: 5.9851034166058525e-05
step: 20, loss: 0.0012422718573361635
step: 30, loss: 4.755987538374029e-05
step: 40, loss: 0.00040744239231571555
step: 50, loss: 0.0017374001909047365
step: 60, loss: 5.0642403948586434e-05
step: 70, loss: 0.0022093551233410835
step: 80, loss: 0.0034638040233403444
step: 90, loss: 0.0015672470908612013
step: 100, loss: 0.00021436730457935482
step: 110, loss: 0.0023861676454544067
step: 120, loss: 0.0012449172791093588
step: 130, loss: 0.00023736804723739624
step: 140, loss: 0.0023508560843765736
step: 150, loss: 0.0003939525340683758
step: 160, loss: 0.00012584814976435155
step: 170, loss: 0.039975401014089584
step: 180, loss: 0.0019263317808508873
step: 190, loss: 0.0010825310600921512
step: 200, loss: 0.005248050671070814
step: 210, loss: 0.0008429366862401366
step: 220, loss: 0.0019499595509842038
step: 230, loss: 0.00030035036616027355
step: 240, loss: 0.00010045323142549023
step: 250, loss: 0.00528772221878171
step: 260, loss: 0.0019126672996208072
step: 270, loss: 0.02097097970545292
step: 280, loss: 0.0006531457765959203
step: 290, loss: 0.029657656326889992
step: 300, loss: 0.0003917836002074182
step: 310, loss: 0.0006890118820592761
step: 320, loss: 0.00015730899758636951
step: 330, loss: 0.0002136011462425813
step: 340, loss: 0.0009835881646722555
step: 350, loss: 0.017641864717006683
step: 360, loss: 0.0005534285446628928
step: 370, loss: 0.005988681688904762
step: 380, loss: 0.018430016934871674
step: 390, loss: 0.00028423446929082274
```

```

step: 400, loss: 0.0033855815418064594
step: 410, loss: 0.00032310641836375
step: 420, loss: 0.04004129022359848
step: 430, loss: 9.252508607460186e-05
step: 440, loss: 0.0004393085837364197
step: 450, loss: 0.01056516170501709
step: 460, loss: 0.013642546720802784
step: 470, loss: 0.00033563465694896877
step: 480, loss: 0.002304369816556573
step: 490, loss: 0.10403550416231155
step: 500, loss: 0.002145844977349043
step: 510, loss: 0.00022383246687240899
step: 520, loss: 0.0012681146617978811
step: 530, loss: 0.00015336929936893284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9354838709677419, f1=0.9350411710887466,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013678941468242556
step: 10, loss: 0.0005189029034227133
step: 20, loss: 8.40374777908437e-05
step: 30, loss: 0.0014058976667001843
step: 40, loss: 0.0020519893150776625
step: 50, loss: 0.001238830154761672
step: 60, loss: 0.0005428397562354803
step: 70, loss: 4.8105237510753796e-05
step: 80, loss: 0.0011948244646191597
step: 90, loss: 0.006082466803491116
step: 100, loss: 8.304600487463176e-05
step: 110, loss: 5.0679303967626765e-05
step: 120, loss: 0.003984002396464348
step: 130, loss: 0.0005271032568998635
step: 140, loss: 0.00014044970157556236
step: 150, loss: 6.572739948751405e-05
step: 160, loss: 0.0031339931301772594
step: 170, loss: 0.00030187010997906327
step: 180, loss: 8.212191460188478e-05
step: 190, loss: 0.00026612880174070597
step: 200, loss: 0.002831097459420562
step: 210, loss: 0.0007057711482048035
step: 220, loss: 0.0011297313030809164
step: 230, loss: 0.024843333289027214
step: 240, loss: 2.655324715306051e-05
step: 250, loss: 0.0006438500131480396

```

```

step: 260, loss: 4.292191079002805e-05
step: 270, loss: 0.0384075865149498
step: 280, loss: 0.0009670339059084654
step: 290, loss: 0.006249593570828438
step: 300, loss: 0.03030138835310936
step: 310, loss: 0.02377997525036335
step: 320, loss: 0.004400694742798805
step: 330, loss: 0.0019942899234592915
step: 340, loss: 4.3680087401298806e-05
step: 350, loss: 0.0042920648120343685
step: 360, loss: 0.00046766723971813917
step: 370, loss: 2.8020996978739277e-05
step: 380, loss: 0.00010121574450749904
step: 390, loss: 0.1655908077955246
step: 400, loss: 6.275817577261478e-05
step: 410, loss: 0.00031100548221729696
step: 420, loss: 0.002134694019332528
step: 430, loss: 0.0008490977925248444
step: 440, loss: 0.0005792936426587403
step: 450, loss: 8.612669626018032e-05
step: 460, loss: 0.001064796349965036
step: 470, loss: 0.007398477289825678
step: 480, loss: 8.743559737922624e-05
step: 490, loss: 0.07849772274494171
step: 500, loss: 0.00012762733967974782
step: 510, loss: 6.18064877926372e-05
step: 520, loss: 3.94654089177493e-05
step: 530, loss: 2.6869247449212708e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9342592592592593, f1=0.9304467987102718,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016506904503330588
step: 10, loss: 6.0190479416633025e-05
step: 20, loss: 2.1073574316687882e-05
step: 30, loss: 0.0002238517627120018
step: 40, loss: 0.033593494445085526
step: 50, loss: 0.04027854651212692
step: 60, loss: 0.00018841531709767878
step: 70, loss: 0.0022720936685800552
step: 80, loss: 0.008637040853500366
step: 90, loss: 0.00983608141541481
step: 100, loss: 5.014482667320408e-05
step: 110, loss: 7.848583481973037e-05

```



```
step: 120, loss: 0.0006369375623762608
step: 130, loss: 7.4264760769438e-05
step: 140, loss: 0.0008921350818127394
step: 150, loss: 0.0013115819310769439
step: 160, loss: 0.0019870023243129253
step: 170, loss: 6.583996582776308e-05
step: 180, loss: 0.007969372905790806
step: 190, loss: 0.0036991622764617205
step: 200, loss: 9.195633174385875e-05
step: 210, loss: 0.00013783694885205477
step: 220, loss: 2.498450703569688e-05
step: 230, loss: 0.0014259510207921267
step: 240, loss: 0.00017922214465215802
step: 250, loss: 0.0009943196782842278
step: 260, loss: 0.00014669103256892413
step: 270, loss: 0.0006791865453124046
step: 280, loss: 0.0015149135142564774
step: 290, loss: 0.0010866884840652347
step: 300, loss: 0.0013050194829702377
step: 310, loss: 0.00020041460811626166
step: 320, loss: 0.0002075753582175821
step: 330, loss: 5.006951323593967e-05
step: 340, loss: 0.0006787400925531983
step: 350, loss: 0.014646850526332855
step: 360, loss: 0.004894144833087921
step: 370, loss: 8.168096974259242e-05
step: 380, loss: 0.0001380718022119254
step: 390, loss: 0.00012328378215897828
step: 400, loss: 0.018859798088669777
step: 410, loss: 0.0001999787345994264
step: 420, loss: 0.0013093568850308657
step: 430, loss: 0.0003484027402009815
step: 440, loss: 0.0007969976286403835
step: 450, loss: 0.0001830641704145819
step: 460, loss: 0.0005231002578511834
step: 470, loss: 0.00013804330956190825
step: 480, loss: 0.0004092518938705325
step: 490, loss: 0.06362784653902054
step: 500, loss: 0.0043398370034992695
step: 510, loss: 7.923805969767272e-05
step: 520, loss: 0.0024985377676784992
step: 530, loss: 0.0004543665563687682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9351724137931035, f1=0.9294492489758761,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0008214824483729899
step: 10, loss: 0.000411155226174742
step: 20, loss: 0.01783776842057705
step: 30, loss: 0.00982519518584013
step: 40, loss: 0.00043474588892422616
step: 50, loss: 0.00037853821413591504
step: 60, loss: 0.00314034684561193
step: 70, loss: 0.0767364427447319
step: 80, loss: 0.00013486384705174714
step: 90, loss: 0.0033851363696157932
step: 100, loss: 0.008843349292874336
step: 110, loss: 0.0017950511537492275
step: 120, loss: 0.0013018575264140964
step: 130, loss: 0.001013843109831214
step: 140, loss: 0.001264907419681549
step: 150, loss: 0.003280829871073365
step: 160, loss: 0.00015713946777395904
step: 170, loss: 0.04766447842121124
step: 180, loss: 8.012907346710563e-05
step: 190, loss: 0.00017542419664096087
step: 200, loss: 0.001494374591857195
step: 210, loss: 0.001913886284455657
step: 220, loss: 0.0006671320297755301
step: 230, loss: 0.04049086943268776
step: 240, loss: 0.0007514652097597718
step: 250, loss: 0.029566537588834763
step: 260, loss: 0.0003572608111426234
step: 270, loss: 0.0026356270536780357
step: 280, loss: 0.0024366045836359262
step: 290, loss: 0.000312742602545768
step: 300, loss: 0.0022895762231200933
step: 310, loss: 0.004558396525681019
step: 320, loss: 0.010891171172261238
step: 330, loss: 0.005324244499206543
step: 340, loss: 0.0010023899376392365
step: 350, loss: 0.031864576041698456
step: 360, loss: 0.00015565271314699203
step: 370, loss: 0.028898106887936592
step: 380, loss: 0.0012449907371774316
step: 390, loss: 0.009325540624558926
step: 400, loss: 0.00025051430566236377
step: 410, loss: 0.0002072157512884587
step: 420, loss: 5.821531158289872e-05
step: 430, loss: 0.05869322642683983
step: 440, loss: 0.0003677466884255409
step: 450, loss: 0.0001779587473720312
```

```

step: 460, loss: 7.848137465771288e-05
step: 470, loss: 0.020957712084054947
step: 480, loss: 0.0002818213542923331
step: 490, loss: 0.0001233920775121078
step: 500, loss: 0.0005277559976093471
step: 510, loss: 0.0004962697275914252
step: 520, loss: 8.142108708852902e-05
step: 530, loss: 0.0005453348276205361
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9368709972552608, f1=0.9338168631006345,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004007955139968544
step: 10, loss: 0.0002444653364364058
step: 20, loss: 0.000711442029569298
step: 30, loss: 8.063200220931321e-05
step: 40, loss: 0.00010850648686755449
step: 50, loss: 0.0010003879433497787
step: 60, loss: 1.4688654118799604e-05
step: 70, loss: 0.00028020134777761996
step: 80, loss: 3.99560340156313e-05
step: 90, loss: 3.4622538805706427e-05
step: 100, loss: 0.0034276910591870546
step: 110, loss: 2.3215377950691618e-05
step: 120, loss: 0.0001842238852987066
step: 130, loss: 3.729783929884434e-05
step: 140, loss: 0.022656267508864403
step: 150, loss: 0.00028838450089097023
step: 160, loss: 0.000862087297718972
step: 170, loss: 0.002985527040436864
step: 180, loss: 7.6026983151678e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 190, loss: 0.00020558973483275622
step: 200, loss: 0.00038203501026146114
step: 210, loss: 0.000158709503011778
step: 220, loss: 0.0002719361800700426
step: 230, loss: 0.00041258649434894323
step: 240, loss: 0.0014033393235877156
step: 250, loss: 0.0018612900748848915
step: 260, loss: 0.0001726097398204729
step: 270, loss: 0.00244273548014462
step: 280, loss: 0.00010842550545930862
step: 290, loss: 0.00039907239261083305
step: 300, loss: 0.00014127849135547876

```

```

step: 310, loss: 1.918120324262418e-05
step: 320, loss: 0.0011077882954850793
step: 330, loss: 5.471596523420885e-05
step: 340, loss: 6.242485687835142e-05
step: 350, loss: 0.006111267954111099
step: 360, loss: 0.008656750433146954
step: 370, loss: 0.0013604846317321062
step: 380, loss: 3.072394974878989e-05
step: 390, loss: 4.01850302296225e-05
step: 400, loss: 6.67992135277018e-05
step: 410, loss: 3.612460204749368e-05
step: 420, loss: 2.4932283849921077e-05
step: 430, loss: 0.000430981075624004
step: 440, loss: 5.7001696404768154e-05
step: 450, loss: 0.0002255524741485715
step: 460, loss: 0.002710256027057767
step: 470, loss: 8.759810589253902e-05
step: 480, loss: 2.2775913748773746e-05
step: 490, loss: 4.201882256893441e-05
step: 500, loss: 0.0002318158803973347
step: 510, loss: 0.0003114034188911319
step: 520, loss: 0.0005857961950823665
step: 530, loss: 0.0031943887006491423
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9367441860465117, f1=0.9391143911439114,
best_f1=0.9323447636700649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005539475823752582
step: 10, loss: 0.0013580179074779153
step: 20, loss: 4.268471820978448e-05
step: 30, loss: 0.00014422026288229972
step: 40, loss: 0.00015663511294405907
step: 50, loss: 0.00011150281352456659
step: 60, loss: 7.026414823485538e-05
step: 70, loss: 0.0010450859554111958
step: 80, loss: 0.00042750893044285476
step: 90, loss: 1.1015619747922756e-05
step: 100, loss: 3.689600998768583e-05
step: 110, loss: 0.009627013467252254
step: 120, loss: 0.0013007413363084197
step: 130, loss: 0.0005299770273268223
step: 140, loss: 0.00012395203521009535
step: 150, loss: 0.0006790108163841069
step: 160, loss: 5.454711936181411e-05

```

```

step: 170, loss: 0.00080193328903988
step: 180, loss: 0.0008534591761417687
step: 190, loss: 0.009217685088515282
step: 200, loss: 0.0009275043266825378
step: 210, loss: 0.00018219189951196313
step: 220, loss: 1.693119338597171e-05
step: 230, loss: 0.02375396527349949
step: 240, loss: 0.0001589375315234065
step: 250, loss: 4.140990131418221e-05
step: 260, loss: 0.0007385708158835769
step: 270, loss: 0.024946516379714012
step: 280, loss: 8.638531289761886e-05
step: 290, loss: 0.00031733501236885786
step: 300, loss: 0.0055544194765388966
step: 310, loss: 5.4410702432505786e-05
step: 320, loss: 8.153488306561485e-05
step: 330, loss: 0.01935115084052086
step: 340, loss: 0.0001764971239026636
step: 350, loss: 8.152816008077934e-05
step: 360, loss: 0.0001946712873177603
step: 370, loss: 0.00011382942466298118
step: 380, loss: 4.5176890125731006e-05
step: 390, loss: 2.961420796054881e-05
step: 400, loss: 0.00030481728026643395
step: 410, loss: 4.6434422984020784e-05
step: 420, loss: 8.184769831132144e-05
step: 430, loss: 2.213476000179071e-05
step: 440, loss: 0.1795879453420639
step: 450, loss: 5.3829611715627834e-05
step: 460, loss: 3.316431320854463e-05
step: 470, loss: 0.0025295123923569918
step: 480, loss: 0.001999207306653261
step: 490, loss: 3.0545917979907244e-05
step: 500, loss: 0.0006876664119772613
step: 510, loss: 0.007819972932338715
step: 520, loss: 0.0005596973351202905
step: 530, loss: 2.091300666506868e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9365808823529411, f1=0.9328460484239378,
best_f1=0.9323447636700649
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled          : True

```

```

opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                  : True
opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:23, 244.81it/s]
load_f1 = 0.9402985074626865
real_f1 = 0.9387186629526463
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:18, 243.69it/s]

```

## 2 BASELINE

### 2.1 BASELINE STRUCTURED

#### 2.1.1 Beer - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Structured/Beer \
        --batch_size 32 \

```

```

--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
 /usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:  
UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.  
In PyTorch 1.1.0 and later, you should call them in the opposite order:  
`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result  
in PyTorch skipping the first value of the learning rate schedule. See more  
details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>  
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",  
UserWarning)  
step: 0, loss: 0.8627287745475769  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:  
UserWarning: Please also save or load the state of the optimizer when saving or  
loading the scheduler.  
warnings.warn(SAVE\_STATE\_WARNING, UserWarning)  
epoch 1: dev\_f1=0.2916666666666667, f1=0.2692307692307693,  
best\_f1=0.2692307692307693  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.3643985688686371  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:  
UserWarning: Please also save or load the state of the optimizer when saving or  
loading the scheduler.  
warnings.warn(SAVE\_STATE\_WARNING, UserWarning)  
epoch 2: dev\_f1=0.39215686274509803, f1=0.3058823529411765,  
best\_f1=0.3058823529411765  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.3405272364616394  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
epoch 3: dev\_f1=0.3846153846153846, f1=0.2857142857142857,  
best\_f1=0.3058823529411765  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.3413766324520111  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.



```

    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.4761904761904762, f1=0.32432432432432434,
best_f1=0.32432432432432434
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22042208909988403
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.5405405405405405, f1=0.32142857142857145,
best_f1=0.32142857142857145
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3019135594367981
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.5806451612903226, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19330568611621857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.6000000000000001, f1=0.4166666666666667,
best_f1=0.4166666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.2851102948188782
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.6250000000000001, f1=0.44897959183673464,
best_f1=0.44897959183673464
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25105515122413635
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.6666666666666665, f1=0.43999999999999995,
best_f1=0.43999999999999995
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1370418220758438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.6857142857142857, f1=0.46511627906976755,
best_f1=0.46511627906976755
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1135677695274353
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.6923076923076924, f1=0.47058823529411764,

```

```

best_f1=0.47058823529411764
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12602001428604126
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.7586206896551724, f1=0.5000000000000001,
best_f1=0.5000000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.035903699696063995
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7586206896551724, f1=0.5000000000000001,
best_f1=0.5000000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06849225610494614
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7586206896551724, f1=0.5142857142857143,
best_f1=0.5000000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08042121678590775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7586206896551724, f1=0.5142857142857143,
best_f1=0.5000000000000001
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02

```

```

cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled              : True
opt_level            : 02
cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
91it [00:00, 134822.21it/s]
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
load_f1 = 0.7333333333333334
real_f1 = 0.7096774193548386
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:16, 270.19it/s]

```

### 2.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \

```

```

--save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more

details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.8118889331817627
```

```
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
```

```
step: 10, loss: 0.4738282263278961
```

```
step: 20, loss: 0.5766174793243408
```

```
step: 30, loss: 0.45039254426956177
```

```
step: 40, loss: 0.2614929974079132
```

```
step: 50, loss: 0.08854293823242188
```

```
step: 60, loss: 0.11620572209358215
```

```
step: 70, loss: 0.08670017868280411
```

```
step: 80, loss: 0.15161927044391632
```

```
step: 90, loss: 0.04919730871915817
```

```
step: 100, loss: 0.017695968970656395
```

```
step: 110, loss: 0.04437314346432686
```

```
step: 120, loss: 0.08302196115255356
```

```
step: 130, loss: 0.054010432213544846
```

```
step: 140, loss: 0.09156283736228943
```

```
step: 150, loss: 0.06887815147638321
```

```
step: 160, loss: 0.15386810898780823
```

```
step: 170, loss: 0.004770820494741201
```

```
step: 180, loss: 0.006976143456995487
```

```
step: 190, loss: 0.05519191920757294
```

```
step: 200, loss: 0.020364223048090935
```

```
step: 210, loss: 0.0035182370338588953
```

```
step: 220, loss: 0.0054263765923678875
```

```
step: 230, loss: 0.006498444825410843
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
```

```
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
```

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

```
epoch 1: dev_f1=0.9843400447427293, f1=0.9796380090497738,
```

```
best_f1=0.9796380090497738
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.005511316005140543
```

```
step: 10, loss: 0.0011346053797751665
```

```
step: 20, loss: 0.0075200824066996574
```

```
step: 30, loss: 0.0017158754635602236
```

```
step: 40, loss: 0.024634405970573425
```

```
step: 50, loss: 0.08229091018438339
```

```
step: 60, loss: 0.0042440881952643394
```

```

step: 70, loss: 0.0038895062170922756
step: 80, loss: 0.019416840746998787
step: 90, loss: 0.029403984546661377
step: 100, loss: 0.0038019123021513224
step: 110, loss: 0.017104126513004303
step: 120, loss: 0.009498551487922668
step: 130, loss: 0.0005992905935272574
step: 140, loss: 0.27039894461631775
step: 150, loss: 0.011202404275536537
step: 160, loss: 0.08677157014608383
step: 170, loss: 0.007994793355464935
step: 180, loss: 0.002586409682407975
step: 190, loss: 0.03305681422352791
step: 200, loss: 0.0019012826960533857
step: 210, loss: 0.0308444295078516
step: 220, loss: 0.0006403551669791341
step: 230, loss: 0.01735972799360752
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9887387387387387, f1=0.9853107344632768,
best_f1=0.9853107344632768
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0067468201741576195
step: 10, loss: 0.014678876847028732
step: 20, loss: 0.0018837368115782738
step: 30, loss: 0.007554054260253906
step: 40, loss: 0.0448906347155571
step: 50, loss: 0.01003443542867899
step: 60, loss: 0.0007638154784217477
step: 70, loss: 0.0007900713826529682
step: 80, loss: 0.01565740257501602
step: 90, loss: 0.0009077116847038269
step: 100, loss: 0.0005077586974948645
step: 110, loss: 0.0045792520977556705
step: 120, loss: 0.0007061146898195148
step: 130, loss: 0.0007316878181882203
step: 140, loss: 0.0006083964835852385
step: 150, loss: 0.0017145383171737194
step: 160, loss: 0.0015389008913189173
step: 170, loss: 0.005079676862806082
step: 180, loss: 0.0008987363544292748

```

```

step: 190, loss: 0.0007706008036620915
step: 200, loss: 0.0010113604366779327
step: 210, loss: 0.025081675499677658
step: 220, loss: 0.0008985086460597813
step: 230, loss: 0.04491061344742775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.988814317673378, f1=0.9776785714285714,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007321916054934263
step: 10, loss: 0.0008713359711691737
step: 20, loss: 0.00029717045254074037
step: 30, loss: 0.00033720926148816943
step: 40, loss: 0.0009073987021110952
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.0023323972709476948
step: 60, loss: 0.0005492407944984734
step: 70, loss: 0.00197030883282423
step: 80, loss: 0.03360167518258095
step: 90, loss: 0.00851611141115427
step: 100, loss: 0.0005826007109135389
step: 110, loss: 0.007918601855635643
step: 120, loss: 0.09564721584320068
step: 130, loss: 0.006401793099939823
step: 140, loss: 0.00037241188692860305
step: 150, loss: 0.0005203267792239785
step: 160, loss: 0.0004137953510507941
step: 170, loss: 0.0024016150273382664
step: 180, loss: 0.02751050516963005
step: 190, loss: 0.0007185940048657358
step: 200, loss: 0.00038436870090663433
step: 210, loss: 0.0002505639859009534
step: 220, loss: 0.0005355823086574674
step: 230, loss: 0.00042385555570945144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9842696629213483, f1=0.9818181818181818,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042512838263064623
step: 10, loss: 0.17966553568840027
step: 20, loss: 0.04197859391570091
step: 30, loss: 0.10746162384748459
step: 40, loss: 0.0014754653675481677
step: 50, loss: 0.001881952048279345
step: 60, loss: 0.0029575990047305822
step: 70, loss: 0.0004822363844141364
step: 80, loss: 0.0008831388549879193
step: 90, loss: 0.0005308450199663639
step: 100, loss: 0.0334569476544857
step: 110, loss: 0.0011075511574745178
step: 120, loss: 0.029662122949957848
step: 130, loss: 0.010144742205739021
step: 140, loss: 0.000832402496598661
step: 150, loss: 0.00035252125235274434
step: 160, loss: 0.0301944799721241
step: 170, loss: 0.1295207291841507
step: 180, loss: 0.007493947632610798
step: 190, loss: 0.0010919749038293958
step: 200, loss: 0.0009103022748604417
step: 210, loss: 0.0013691415078938007
step: 220, loss: 0.003914590924978256
step: 230, loss: 0.00033788615837693214
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9832026875699889, f1=0.9841628959276018,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004177536349743605
step: 10, loss: 0.0004737957497127354
step: 20, loss: 0.0007987392600625753
step: 30, loss: 0.0003614200104493648
step: 40, loss: 0.0016222636913880706
step: 50, loss: 0.049588754773139954
step: 60, loss: 0.025747938081622124
step: 70, loss: 0.01885293796658516
step: 80, loss: 0.00440248055383563
step: 90, loss: 0.006214052438735962
step: 100, loss: 0.0003282985126134008
step: 110, loss: 0.07733436673879623
step: 120, loss: 0.0004898371989838779
step: 130, loss: 0.005612088367342949

```

```

step: 140, loss: 0.028689078986644745
step: 150, loss: 0.0004498728958424181
step: 160, loss: 0.0005415345658548176
step: 170, loss: 0.00015596032608300447
step: 180, loss: 0.00031365558970719576
step: 190, loss: 0.005873606074601412
step: 200, loss: 0.00031362802837975323
step: 210, loss: 0.00027019783738069236
step: 220, loss: 0.00022332755906973034
step: 230, loss: 0.00015141494804993272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9853438556933484, f1=0.9805714285714285,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05582502856850624
step: 10, loss: 0.00029748756787739694
step: 20, loss: 0.010363101959228516
step: 30, loss: 0.00027276051696389914
step: 40, loss: 0.09860117733478546
step: 50, loss: 0.00012476198025979102
step: 60, loss: 0.01380450464785099
step: 70, loss: 0.00016787057393230498
step: 80, loss: 0.001055483240634203
step: 90, loss: 0.0008393774041905999
step: 100, loss: 0.00386839103884995
step: 110, loss: 0.007124989293515682
step: 120, loss: 0.0002659049059730023
step: 130, loss: 0.00028976850444450974
step: 140, loss: 0.00019867508672177792
step: 150, loss: 0.0002476218796800822
step: 160, loss: 0.00024891248904168606
step: 170, loss: 0.0004833014099858701
step: 180, loss: 0.0002801259979605675
step: 190, loss: 0.0004934041062369943
step: 200, loss: 0.0006266592536121607
step: 210, loss: 0.0002593879762571305
step: 220, loss: 0.0009298957302235067
step: 230, loss: 0.02608032524585724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9841986455981941, f1=0.9771689497716894,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02724502980709076
step: 10, loss: 0.0016592518659308553
step: 20, loss: 0.0020010238513350487
step: 30, loss: 0.00026727630756795406
step: 40, loss: 0.00013922745711170137
step: 50, loss: 0.00022762990556657314
step: 60, loss: 0.00012332208279985934
step: 70, loss: 0.0006455348338931799
step: 80, loss: 0.00024686421966180205
step: 90, loss: 0.00025558500783517957
step: 100, loss: 0.0002409285953035578
step: 110, loss: 0.0001119152657338418
step: 120, loss: 9.600924386177212e-05
step: 130, loss: 0.13617348670959473
step: 140, loss: 0.00033583244658075273
step: 150, loss: 0.000520881381817162
step: 160, loss: 0.00021395954536274076
step: 170, loss: 0.0014890307793393731
step: 180, loss: 0.0009257674682885408
step: 190, loss: 0.00021286342234816402
step: 200, loss: 0.003359259106218815
step: 210, loss: 0.0001511375594418496
step: 220, loss: 0.00012582039926201105
step: 230, loss: 0.01170478854328394
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9853438556933484, f1=0.9794988610478361,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005016321665607393
step: 10, loss: 0.00010302464943379164
step: 20, loss: 0.00025047530652955174
step: 30, loss: 0.00011138302943436429
step: 40, loss: 0.00041812710696831346
step: 50, loss: 6.035925980540924e-05
step: 60, loss: 0.00010890594421653077
step: 70, loss: 0.00016238601529039443
step: 80, loss: 0.02655557543039322
step: 90, loss: 0.0011843611719086766
step: 100, loss: 0.00011767962132580578
step: 110, loss: 0.0005395112675614655
step: 120, loss: 0.030266735702753067
step: 130, loss: 7.429585821228102e-05

```

```

step: 140, loss: 0.0001336300774710253
step: 150, loss: 5.952649371465668e-05
step: 160, loss: 0.0001493588788434863
step: 170, loss: 4.996137795387767e-05
step: 180, loss: 0.00012704548134934157
step: 190, loss: 0.00017146825848612934
step: 200, loss: 0.0001184953362098895
step: 210, loss: 0.00010280866990797222
step: 220, loss: 0.027478432282805443
step: 230, loss: 0.01111991424113512
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9864253393665158, f1=0.9794050343249429,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017287499504163861
step: 10, loss: 0.0012453461531549692
step: 20, loss: 0.0001310354855377227
step: 30, loss: 0.0001545120176160708
step: 40, loss: 0.00013385873171500862
step: 50, loss: 0.10206151008605957
step: 60, loss: 0.02540714293718338
step: 70, loss: 0.011914944276213646
step: 80, loss: 0.00021341034153010696
step: 90, loss: 0.00011468524462543428
step: 100, loss: 9.015985415317118e-05
step: 110, loss: 0.00013908755499869585
step: 120, loss: 0.019907040521502495
step: 130, loss: 0.001063030562363565
step: 140, loss: 0.027458375319838524
step: 150, loss: 7.840661419322714e-05
step: 160, loss: 0.00015631478163413703
step: 170, loss: 0.00019092012371402234
step: 180, loss: 0.00018408785399515182
step: 190, loss: 0.00011494107457110658
step: 200, loss: 0.0001139907690230757
step: 210, loss: 8.211396198021248e-05
step: 220, loss: 0.001400155364535749
step: 230, loss: 8.998007979243994e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9853438556933484, f1=0.9806157354618015,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.257662946358323e-05
step: 10, loss: 8.885493298294023e-05
step: 20, loss: 6.448049680329859e-05
step: 30, loss: 0.0008579600835219026
step: 40, loss: 0.003358118236064911
step: 50, loss: 0.02034865692257881
step: 60, loss: 8.81713567650877e-05
step: 70, loss: 0.0001343349285889417
step: 80, loss: 0.00010255639062961563
step: 90, loss: 5.7444362028036267e-05
step: 100, loss: 6.57993441564031e-05
step: 110, loss: 0.00010087454575113952
step: 120, loss: 0.00010881429625442252
step: 130, loss: 0.00010297726839780807
step: 140, loss: 5.856455027242191e-05
step: 150, loss: 4.4813517888542265e-05
step: 160, loss: 0.0002581898879725486
step: 170, loss: 0.007884232327342033
step: 180, loss: 0.0002995034446939826
step: 190, loss: 0.1284344345331192
step: 200, loss: 0.00011642324534477666
step: 210, loss: 7.968972204253078e-05
step: 220, loss: 6.864373426651582e-05
step: 230, loss: 0.022844985127449036
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9842696629213483, f1=0.9806157354618015,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.794288780540228e-05
step: 10, loss: 0.00011403149255784228
step: 20, loss: 8.019310917006806e-05
step: 30, loss: 0.008872786536812782
step: 40, loss: 5.922230411670171e-05
step: 50, loss: 0.00010615258361212909
step: 60, loss: 0.019027117639780045
step: 70, loss: 0.00016297606634907424
step: 80, loss: 0.007986620999872684
step: 90, loss: 8.308881660923362e-05
step: 100, loss: 9.266052802558988e-05
step: 110, loss: 0.00045083340955898166
step: 120, loss: 0.00011435778287705034
step: 130, loss: 0.0006950516253709793
```

```

step: 140, loss: 8.673956472193822e-05
step: 150, loss: 8.713347051525488e-05
step: 160, loss: 0.040986884385347366
step: 170, loss: 0.0001523703249404207
step: 180, loss: 7.542310777353123e-05
step: 190, loss: 0.00011695801367750391
step: 200, loss: 0.01142908725887537
step: 210, loss: 6.70679728500545e-05
step: 220, loss: 6.745233258698136e-05
step: 230, loss: 8.833668834995478e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9853768278965129, f1=0.9817767653758542,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001839546748669818
step: 10, loss: 9.748402226250619e-05
step: 20, loss: 9.019442222779617e-05
step: 30, loss: 0.0688929334282875
step: 40, loss: 8.076774975052103e-05
step: 50, loss: 6.478642171714455e-05
step: 60, loss: 8.281993359560147e-05
step: 70, loss: 6.674041651422158e-05
step: 80, loss: 9.753745689522475e-05
step: 90, loss: 3.9661455957684666e-05
step: 100, loss: 0.00010303875751560554
step: 110, loss: 0.00017308043607044965
step: 120, loss: 0.0001927782577695325
step: 130, loss: 8.48273848532699e-05
step: 140, loss: 8.395413897233084e-05
step: 150, loss: 0.018694080412387848
step: 160, loss: 4.917680053040385e-05
step: 170, loss: 6.221699732122943e-05
step: 180, loss: 0.00014557348913513124
step: 190, loss: 3.741513501154259e-05
step: 200, loss: 0.00012172628339612857
step: 210, loss: 0.0009193583973683417
step: 220, loss: 4.9187790864380077e-05
step: 230, loss: 4.3724958231905475e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9853768278965129, f1=0.9807037457434733,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000776369241066277
step: 10, loss: 2.9402755899354815e-05
step: 20, loss: 0.003161307657137513
step: 30, loss: 0.0001607039594091475
step: 40, loss: 4.3400137656135485e-05
step: 50, loss: 6.804181612096727e-05
step: 60, loss: 5.2699397201649845e-05
step: 70, loss: 6.294017657637596e-05
step: 80, loss: 0.002044226974248886
step: 90, loss: 6.705256237182766e-05
step: 100, loss: 0.012023206800222397
step: 110, loss: 0.0001507754495833069
step: 120, loss: 7.765877671772614e-05
step: 130, loss: 0.00020693274564109743
step: 140, loss: 0.0001251552312169224
step: 150, loss: 0.00012328092998359352
step: 160, loss: 3.8107224099803716e-05
step: 170, loss: 0.0001356971770292148
step: 180, loss: 8.995852840598673e-05
step: 190, loss: 0.007297143340110779
step: 200, loss: 4.5140801375964656e-05
step: 210, loss: 0.017462899908423424
step: 220, loss: 6.318198575172573e-05
step: 230, loss: 2.8065722290193662e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9853768278965129, f1=0.9807037457434733,
best_f1=0.9776785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.405995787237771e-05
step: 10, loss: 0.0001222657592734322
step: 20, loss: 5.617622809950262e-05
step: 30, loss: 0.00016569920990150422
step: 40, loss: 0.00021150201791897416
step: 50, loss: 0.00012570773833431304
step: 60, loss: 5.891818000236526e-05
step: 70, loss: 8.10440833447501e-05
step: 80, loss: 0.00026400561910122633
step: 90, loss: 9.734199556987733e-05
step: 100, loss: 5.9101683291373774e-05
step: 110, loss: 9.877697448246181e-05
step: 120, loss: 0.00013239687541499734
step: 130, loss: 7.533319876529276e-05

```

```

step: 140, loss: 0.00021110961097292602
step: 150, loss: 0.0001760355953592807
step: 160, loss: 7.839257887098938e-05
step: 170, loss: 3.9508235204266384e-05
step: 180, loss: 6.694855255773291e-05
step: 190, loss: 6.358773680403829e-05
step: 200, loss: 4.592649202095345e-05
step: 210, loss: 0.003241589991375804
step: 220, loss: 0.00012680944928433746
step: 230, loss: 0.00017005081463139504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9853768278965129, f1=0.9818594104308391,
best_f1=0.9776785714285714
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 242.10it/s]
load_f1 = 0.9876265466816648
real_f1 = 0.9876265466816648

```



```
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:16, 267.09it/s]
```

### 2.1.3 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
```

```

loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7917187213897705
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45270735025405884
step: 20, loss: 0.49639686942100525
step: 30, loss: 0.4561195969581604
step: 40, loss: 0.4177701771259308
step: 50, loss: 0.22395043075084686
step: 60, loss: 0.16186818480491638
step: 70, loss: 0.08552682399749756
step: 80, loss: 0.09186184406280518
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.15897631645202637
step: 100, loss: 0.37582606077194214
step: 110, loss: 0.05420492962002754
step: 120, loss: 0.07634123414754868
step: 130, loss: 0.02924710139632225
step: 140, loss: 0.30979081988334656
step: 150, loss: 0.042966246604919434
step: 160, loss: 0.11274416744709015
step: 170, loss: 0.1876501888036728
step: 180, loss: 0.1631079614162445
step: 190, loss: 0.03716978803277016

```

```

step: 200, loss: 0.15250147879123688
step: 210, loss: 0.09789654612541199
step: 220, loss: 0.03909801319241524
step: 230, loss: 0.13664814829826355
step: 240, loss: 0.07677634060382843
step: 250, loss: 0.062458351254463196
step: 260, loss: 0.01645580865442753
step: 270, loss: 0.005417660344392061
step: 280, loss: 0.13479967415332794
step: 290, loss: 0.1252540647983551
step: 300, loss: 0.08998699486255646
step: 310, loss: 0.04877594858407974
step: 320, loss: 0.07605293393135071
step: 330, loss: 0.10858464986085892
step: 340, loss: 0.16920872032642365
step: 350, loss: 0.022157693281769753
step: 360, loss: 0.09356548637151718
step: 370, loss: 0.14053502678871155
step: 380, loss: 0.17450591921806335
step: 390, loss: 0.016977354884147644
step: 400, loss: 0.006757013034075499
step: 410, loss: 0.06209053099155426
step: 420, loss: 0.018332868814468384
step: 430, loss: 0.03506780415773392
step: 440, loss: 0.072883240878582
step: 450, loss: 0.059737369418144226
step: 460, loss: 0.18977341055870056
step: 470, loss: 0.24054008722305298
step: 480, loss: 0.3045846223831177
step: 490, loss: 0.04008197411894798
step: 500, loss: 0.027106601744890213
step: 510, loss: 0.11368750780820847
step: 520, loss: 0.030016744509339333
step: 530, loss: 0.1019393652677536
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9330872173511767, f1=0.9248501613646842,
best_f1=0.9248501613646842
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1055079996585846
step: 10, loss: 0.13042126595973969

```

step: 20, loss: 0.1142532229423523  
step: 30, loss: 0.04222056642174721  
step: 40, loss: 0.005287847016006708  
step: 50, loss: 0.08448950946331024  
step: 60, loss: 0.17382951080799103  
step: 70, loss: 0.0832710936665535  
step: 80, loss: 0.005449483636766672  
step: 90, loss: 0.013727426528930664  
step: 100, loss: 0.40022867918014526  
step: 110, loss: 0.011915978975594044  
step: 120, loss: 0.08393227308988571  
step: 130, loss: 0.05550137162208557  
step: 140, loss: 0.021936265751719475  
step: 150, loss: 0.04922567307949066  
step: 160, loss: 0.017056120559573174  
step: 170, loss: 0.2101697325706482  
step: 180, loss: 0.01633146032691002  
step: 190, loss: 0.028227459639310837  
step: 200, loss: 0.01130551379173994  
step: 210, loss: 0.006831763312220573  
step: 220, loss: 0.1271182894706726  
step: 230, loss: 0.012176837772130966  
step: 240, loss: 0.17564333975315094  
step: 250, loss: 0.019206229597330093  
step: 260, loss: 0.007469235919415951  
step: 270, loss: 0.11061392724514008  
step: 280, loss: 0.39221781492233276  
step: 290, loss: 0.12114519625902176  
step: 300, loss: 0.016520606353878975  
step: 310, loss: 0.04507740959525108  
step: 320, loss: 0.13931488990783691  
step: 330, loss: 0.04561571404337883  
step: 340, loss: 0.08122394979000092  
step: 350, loss: 0.037959568202495575  
step: 360, loss: 0.04296493902802467  
step: 370, loss: 0.0033217095769941807  
step: 380, loss: 0.04441811889410019  
step: 390, loss: 0.028064727783203125  
step: 400, loss: 0.032941095530986786  
step: 410, loss: 0.001268615829758346  
step: 420, loss: 0.029140489175915718  
step: 430, loss: 0.0287301205098629  
step: 440, loss: 0.04373359680175781  
step: 450, loss: 0.016228528693318367  
step: 460, loss: 0.1632138192653656  
step: 470, loss: 0.1290280967950821  
step: 480, loss: 0.25307556986808777  
step: 490, loss: 0.03865726664662361

```

step: 500, loss: 0.034818731248378754
step: 510, loss: 0.09806978702545166
step: 520, loss: 0.10622797906398773
step: 530, loss: 0.15172800421714783
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.9317972350230416, f1=0.9340710004610421,
best_f1=0.9248501613646842
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030301077291369438
step: 10, loss: 0.08565317839384079
step: 20, loss: 0.135570228099823
step: 30, loss: 0.19018736481666565
step: 40, loss: 0.01046841312199831
step: 50, loss: 0.02106023021042347
step: 60, loss: 0.004690456669777632
step: 70, loss: 0.03954753652215004
step: 80, loss: 0.0018085988704115152
step: 90, loss: 0.022635530680418015
step: 100, loss: 0.07794693112373352
step: 110, loss: 0.030949894338846207
step: 120, loss: 0.053961820900440216
step: 130, loss: 0.030971145257353783
step: 140, loss: 0.006395620759576559
step: 150, loss: 0.01744496263563633
step: 160, loss: 0.0033024889416992664
step: 170, loss: 0.004396950826048851
step: 180, loss: 0.0022972554434090853
step: 190, loss: 0.0008039182284846902
step: 200, loss: 0.02127663977444172
step: 210, loss: 0.03465575724840164
step: 220, loss: 0.029948918148875237
step: 230, loss: 0.030190249904990196
step: 240, loss: 0.026875508949160576
step: 250, loss: 0.008240007795393467
step: 260, loss: 0.0024691340513527393
step: 270, loss: 0.001681128516793251
step: 280, loss: 0.0016447261441498995
step: 290, loss: 0.11492431163787842
step: 300, loss: 0.016907211393117905
step: 310, loss: 0.08520536124706268
step: 320, loss: 0.08221153169870377
step: 330, loss: 0.0020961719565093517
step: 340, loss: 0.0029858993366360664
step: 350, loss: 0.01287498977035284

```

```

step: 360, loss: 0.03339005261659622
step: 370, loss: 0.0022717793472111225
step: 380, loss: 0.027927542105317116
step: 390, loss: 0.015318932943046093
step: 400, loss: 0.071950763463974
step: 410, loss: 0.10419628769159317
step: 420, loss: 0.06542504578828812
step: 430, loss: 0.02925020642578602
step: 440, loss: 0.018912145867943764
step: 450, loss: 0.05248217657208443
step: 460, loss: 0.10024598985910416
step: 470, loss: 0.003760176245123148
step: 480, loss: 0.010203100740909576
step: 490, loss: 0.012690753675997257
step: 500, loss: 0.036211512982845306
step: 510, loss: 0.023530589416623116
step: 520, loss: 0.018175654113292694
step: 530, loss: 0.05209044739603996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9343807763401109, f1=0.9337626494940202,
best_f1=0.9337626494940202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006937011610716581
step: 10, loss: 0.006424611434340477
step: 20, loss: 0.037928931415081024
step: 30, loss: 0.010468262247741222
step: 40, loss: 0.013140557333827019
step: 50, loss: 0.007606544531881809
step: 60, loss: 0.0006419467972591519
step: 70, loss: 0.0004810671671293676
step: 80, loss: 0.01351966056972742
step: 90, loss: 0.019863111898303032
step: 100, loss: 0.01101798377931118
step: 110, loss: 0.0038688508793711662
step: 120, loss: 0.006736171431839466
step: 130, loss: 0.002346575492992997
step: 140, loss: 0.01205662451684475
step: 150, loss: 0.0011580850696191192
step: 160, loss: 0.001207080204039812
step: 170, loss: 0.0027785194106400013

```

```

step: 180, loss: 0.005659324117004871
step: 190, loss: 0.0223334189504385
step: 200, loss: 0.0015041778096929193
step: 210, loss: 0.0540878027677536
step: 220, loss: 0.02617398276925087
step: 230, loss: 0.17966173589229584
step: 240, loss: 0.019661666825413704
step: 250, loss: 0.0013917643809691072
step: 260, loss: 0.08699385821819305
step: 270, loss: 0.02421366237103939
step: 280, loss: 0.002146216342225671
step: 290, loss: 0.020161181688308716
step: 300, loss: 0.0005243784398771822
step: 310, loss: 0.007017194759100676
step: 320, loss: 0.056223224848508835
step: 330, loss: 0.05649961903691292
step: 340, loss: 0.019492607563734055
step: 350, loss: 0.023923667147755623
step: 360, loss: 0.010018492117524147
step: 370, loss: 0.08546872437000275
step: 380, loss: 0.006402689032256603
step: 390, loss: 0.023149438202381134
step: 400, loss: 0.007443887181580067
step: 410, loss: 0.0015329086454585195
step: 420, loss: 0.004345856606960297
step: 430, loss: 0.009753827936947346
step: 440, loss: 0.20515283942222595
step: 450, loss: 0.024969566613435745
step: 460, loss: 0.009945526719093323
step: 470, loss: 0.005832873750478029
step: 480, loss: 0.0031763343140482903
step: 490, loss: 0.026343457400798798
step: 500, loss: 0.009626359678804874
step: 510, loss: 0.07742570340633392
step: 520, loss: 0.006140316836535931
step: 530, loss: 0.0016617312794551253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9352319706017456, f1=0.930939226519337,
best_f1=0.930939226519337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.005076419096440077  
step: 10, loss: 0.00901535339653492  
step: 20, loss: 0.00045970070641487837  
step: 30, loss: 0.00044664705637842417  
step: 40, loss: 0.02366219088435173  
step: 50, loss: 0.019467875361442566  
step: 60, loss: 0.002651932882145047  
step: 70, loss: 0.0009607455576770008  
step: 80, loss: 0.0006220408831723034  
step: 90, loss: 0.028802428394556046  
step: 100, loss: 0.00032755013671703637  
step: 110, loss: 0.011011089198291302  
step: 120, loss: 0.003545035608112812  
step: 130, loss: 0.00027019946719519794  
step: 140, loss: 0.04972009360790253  
step: 150, loss: 0.0021025571040809155  
step: 160, loss: 0.17575043439865112  
step: 170, loss: 0.012068399228155613  
step: 180, loss: 0.1026666983962059  
step: 190, loss: 0.0006632930017076433  
step: 200, loss: 0.0015271937008947134  
step: 210, loss: 0.005216287449002266  
step: 220, loss: 0.0005645535420626402  
step: 230, loss: 0.0007725066388957202  
step: 240, loss: 0.0050346567295491695  
step: 250, loss: 0.0004523429088294506  
step: 260, loss: 0.02171546407043934  
step: 270, loss: 0.017369907349348068  
step: 280, loss: 0.0437040701508522  
step: 290, loss: 0.0017248363001272082  
step: 300, loss: 0.024606676772236824  
step: 310, loss: 0.0014753827126696706  
step: 320, loss: 0.014248911291360855  
step: 330, loss: 0.04852592945098877  
step: 340, loss: 0.004137046635150909  
step: 350, loss: 0.053055062890052795  
step: 360, loss: 0.013692371547222137  
step: 370, loss: 0.001078597386367619  
step: 380, loss: 0.10392820090055466  
step: 390, loss: 0.002396448515355587  
step: 400, loss: 0.043059639632701874  
step: 410, loss: 0.0003924517659470439  
step: 420, loss: 0.006746363360434771  
step: 430, loss: 0.006394997239112854  
step: 440, loss: 0.002316309604793787  
step: 450, loss: 0.02660195343196392  
step: 460, loss: 0.0020481636747717857  
step: 470, loss: 0.0057405452243983746



```

step: 480, loss: 0.0022363353054970503
step: 490, loss: 0.0036781099624931812
step: 500, loss: 0.05454500392079353
step: 510, loss: 0.06669285893440247
step: 520, loss: 0.0009666221449151635
step: 530, loss: 0.005515322554856539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9338919925512105, f1=0.9324639031206335,
best_f1=0.930939226519337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002098131226375699
step: 10, loss: 0.000401430472265929
step: 20, loss: 0.001982511021196842
step: 30, loss: 0.0020631286315619946
step: 40, loss: 0.007283344864845276
step: 50, loss: 0.01085479836910963
step: 60, loss: 0.001543713384307921
step: 70, loss: 0.0020289986860007048
step: 80, loss: 0.0010547382989898324
step: 90, loss: 0.0006195163587108254
step: 100, loss: 0.12068350613117218
step: 110, loss: 0.1315578669309616
step: 120, loss: 0.0027916922699660063
step: 130, loss: 0.002238371642306447
step: 140, loss: 0.02255439944565296
step: 150, loss: 0.006363177206367254
step: 160, loss: 0.05133375898003578
step: 170, loss: 0.1395917534828186
step: 180, loss: 0.058011773973703384
step: 190, loss: 0.030579984188079834
step: 200, loss: 0.00045195568236522377
step: 210, loss: 0.001686609466560185
step: 220, loss: 0.0011176548432558775
step: 230, loss: 0.0030633811838924885
step: 240, loss: 0.001361596048809588
step: 250, loss: 0.000197846646187827
step: 260, loss: 0.0011235070414841175
step: 270, loss: 0.005901236552745104
step: 280, loss: 0.004074288997799158
step: 290, loss: 0.005222305655479431
step: 300, loss: 0.0007225649314932525
step: 310, loss: 0.0003213193849660456
step: 320, loss: 0.08334111422300339
step: 330, loss: 0.0044573028571903706

```

```

step: 340, loss: 0.001570007181726396
step: 350, loss: 0.08184010535478592
step: 360, loss: 0.009325982071459293
step: 370, loss: 0.00041106482967734337
step: 380, loss: 0.08297009021043777
step: 390, loss: 0.003027163213118911
step: 400, loss: 0.03424495831131935
step: 410, loss: 0.0015409099869430065
step: 420, loss: 0.007811833638697863
step: 430, loss: 0.0002552336663939059
step: 440, loss: 0.08260791003704071
step: 450, loss: 0.00188393983989954
step: 460, loss: 0.04036659747362137
step: 470, loss: 0.009221031330525875
step: 480, loss: 0.006367187947034836
step: 490, loss: 0.00016577716451138258
step: 500, loss: 0.01952156238257885
step: 510, loss: 0.0221481341868639
step: 520, loss: 0.009088424034416676
step: 530, loss: 0.002618596889078617
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.937995337995338, f1=0.9358914365933552,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001940157264471054
step: 10, loss: 0.01137796975672245
step: 20, loss: 0.004571973346173763
step: 30, loss: 0.006392620503902435
step: 40, loss: 0.0003451933152973652
step: 50, loss: 0.002670921152457595
step: 60, loss: 0.0004638584505300969
step: 70, loss: 0.003762627951800823
step: 80, loss: 0.0002590447838883847
step: 90, loss: 0.00014854159962851554
step: 100, loss: 0.00022192786855157465
step: 110, loss: 0.0008402051171287894
step: 120, loss: 0.000897709047421813
step: 130, loss: 0.011844965629279613
step: 140, loss: 0.001637935172766447
step: 150, loss: 0.00012714268814306706

```

```

step: 160, loss: 0.003529856214299798
step: 170, loss: 0.0006813023355789483
step: 180, loss: 0.0004497975460253656
step: 190, loss: 0.0006891558296047151
step: 200, loss: 0.0013283746084198356
step: 210, loss: 0.0006744041456840932
step: 220, loss: 0.00014729404938407242
step: 230, loss: 0.00014185329200699925
step: 240, loss: 0.0025995969772338867
step: 250, loss: 0.0029231656808406115
step: 260, loss: 0.004299584776163101
step: 270, loss: 0.0017812576843425632
step: 280, loss: 0.03211018443107605
step: 290, loss: 0.003869433421641588
step: 300, loss: 0.0008523218566551805
step: 310, loss: 0.00608487194404006
step: 320, loss: 0.06100992485880852
step: 330, loss: 7.215829100459814e-05
step: 340, loss: 0.004592386540025473
step: 350, loss: 0.0011576717952266335
step: 360, loss: 0.004124699160456657
step: 370, loss: 0.00020750841940753162
step: 380, loss: 0.001693498925305903
step: 390, loss: 0.00019707821775227785
step: 400, loss: 0.00022078996698837727
step: 410, loss: 0.008366692811250687
step: 420, loss: 0.0006374387303367257
step: 430, loss: 6.568760727532208e-05
step: 440, loss: 0.0004200766852591187
step: 450, loss: 0.01746176928281784
step: 460, loss: 0.002034219214692712
step: 470, loss: 0.0013741585426032543
step: 480, loss: 0.05113597214221954
step: 490, loss: 7.242634455906227e-05
step: 500, loss: 5.927680103923194e-05
step: 510, loss: 0.0011019870871677995
step: 520, loss: 0.0002407282154308632
step: 530, loss: 0.0004381824692245573
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9333958724202628, f1=0.9338959212376934,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01012807060033083
step: 10, loss: 0.006256883032619953

```

step: 20, loss: 0.016241958364844322  
step: 30, loss: 0.011954071931540966  
step: 40, loss: 0.002057202160358429  
step: 50, loss: 0.003269194858148694  
step: 60, loss: 8.247051300713792e-05  
step: 70, loss: 0.00011424689728301018  
step: 80, loss: 0.00014412036398425698  
step: 90, loss: 0.005840519908815622  
step: 100, loss: 0.017288928851485252  
step: 110, loss: 0.0011159561108797789  
step: 120, loss: 0.004176326561719179  
step: 130, loss: 0.00032855928293429315  
step: 140, loss: 5.139729910297319e-05  
step: 150, loss: 0.0004310225776862353  
step: 160, loss: 0.008887429721653461  
step: 170, loss: 0.000115762377390638  
step: 180, loss: 0.00011280018225079402  
step: 190, loss: 0.0002879880485124886  
step: 200, loss: 0.0023360121995210648  
step: 210, loss: 0.03327340632677078  
step: 220, loss: 0.0010065591195598245  
step: 230, loss: 0.00010492724686628208  
step: 240, loss: 0.00027752824826166034  
step: 250, loss: 0.0006621345528401434  
step: 260, loss: 0.007466403767466545  
step: 270, loss: 0.00010156686039408669  
step: 280, loss: 0.0006090649403631687  
step: 290, loss: 0.0028616427443921566  
step: 300, loss: 3.9682345231994987e-05  
step: 310, loss: 0.028776243329048157  
step: 320, loss: 0.0009267692221328616  
step: 330, loss: 0.003494387026876211  
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0  
step: 340, loss: 0.0003900528245139867  
step: 350, loss: 0.009882815182209015  
step: 360, loss: 0.002509861718863249  
step: 370, loss: 0.00010469659900991246  
step: 380, loss: 0.0024126428179442883  
step: 390, loss: 0.01257841382175684  
step: 400, loss: 0.14518007636070251  
step: 410, loss: 4.481976793613285e-05  
step: 420, loss: 9.250619041267782e-05  
step: 430, loss: 0.00011240697494940832  
step: 440, loss: 0.000262944377027452  
step: 450, loss: 0.0012930432567372918  
step: 460, loss: 0.00014319553156383336  
step: 470, loss: 0.0007532263407483697  
step: 480, loss: 0.00011098141112597659

```

step: 490, loss: 6.0118338296888396e-05
step: 500, loss: 5.2118277380941436e-05
step: 510, loss: 0.00021331764583010226
step: 520, loss: 0.00012353886268101633
step: 530, loss: 4.713453017757274e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9368520263901979, f1=0.9316360207449317,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.251284164842218e-05
step: 10, loss: 0.00040945393266156316
step: 20, loss: 5.152074299985543e-05
step: 30, loss: 6.392227078322321e-05
step: 40, loss: 0.00020516468794085085
step: 50, loss: 4.954258110956289e-05
step: 60, loss: 5.326477548805997e-05
step: 70, loss: 0.13735869526863098
step: 80, loss: 4.839591201744042e-05
step: 90, loss: 3.586900857044384e-05
step: 100, loss: 8.16346873762086e-05
step: 110, loss: 4.094944233656861e-05
step: 120, loss: 4.021217318950221e-05
step: 130, loss: 4.1710096411406994e-05
step: 140, loss: 0.00110700074583292
step: 150, loss: 0.00032734329579398036
step: 160, loss: 4.645809531211853e-05
step: 170, loss: 0.0007739452412351966
step: 180, loss: 3.2147858291864395e-05
step: 190, loss: 0.003174275392666459
step: 200, loss: 0.000443979719420895
step: 210, loss: 0.0009026728221215308
step: 220, loss: 0.05060489475727081
step: 230, loss: 0.00016514728486072272
step: 240, loss: 0.0003710499149747193
step: 250, loss: 0.0004525921249296516
step: 260, loss: 0.0018885850440710783
step: 270, loss: 0.0005438460502773523
step: 280, loss: 3.716193896252662e-05
step: 290, loss: 5.1137933041900396e-05
step: 300, loss: 5.636990681523457e-05
step: 310, loss: 6.368570029735565e-05
step: 320, loss: 0.00018262291268911213
step: 330, loss: 0.0017470623133704066
step: 340, loss: 3.390153142390773e-05

```

```

step: 350, loss: 3.396185638848692e-05
step: 360, loss: 0.02743532322347164
step: 370, loss: 0.00010480171476956457
step: 380, loss: 4.0010709199123085e-05
step: 390, loss: 0.0001559888623887673
step: 400, loss: 0.0001422794011887163
step: 410, loss: 5.5614207667531446e-05
step: 420, loss: 0.018718648701906204
step: 430, loss: 3.145485607092269e-05
step: 440, loss: 4.7677265683887526e-05
step: 450, loss: 3.351896884851158e-05
step: 460, loss: 9.831894567469135e-05
step: 470, loss: 7.648350583622232e-05
step: 480, loss: 0.00014202939928509295
step: 490, loss: 6.156459858175367e-05
step: 500, loss: 0.0013487893156707287
step: 510, loss: 0.014704901725053787
step: 520, loss: 0.007237217389047146
step: 530, loss: 0.00037769947084598243
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9330166270783848, f1=0.9321953532479849,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006337047670967877
step: 10, loss: 1.930748658196535e-05
step: 20, loss: 0.0002206883509643376
step: 30, loss: 0.0008287120726890862
step: 40, loss: 0.0002867389703169465
step: 50, loss: 0.0020243225153535604
step: 60, loss: 0.00015358638484030962
step: 70, loss: 8.615561091573909e-05
step: 80, loss: 0.07703576982021332
step: 90, loss: 0.0006153200520202518
step: 100, loss: 0.0002636183344293386
step: 110, loss: 0.001327616279013455
step: 120, loss: 0.00015422045544255525
step: 130, loss: 6.821237184340134e-05
step: 140, loss: 0.0016847150400280952
step: 150, loss: 3.928432852262631e-05
step: 160, loss: 0.00022422130859922618
step: 170, loss: 0.00015302543761208653
step: 180, loss: 0.0002455018984619528
step: 190, loss: 0.0002403332618996501
step: 200, loss: 0.0014703626511618495

```

```

step: 210, loss: 0.0003684069961309433
step: 220, loss: 7.370852108579129e-05
step: 230, loss: 0.00046580753405578434
step: 240, loss: 3.928128717234358e-05
step: 250, loss: 0.00026512466138228774
step: 260, loss: 0.021599136292934418
step: 270, loss: 0.0011360542848706245
step: 280, loss: 0.000257220322964713
step: 290, loss: 5.807447087136097e-05
step: 300, loss: 0.00035513777402229607
step: 310, loss: 7.047885446809232e-05
step: 320, loss: 0.0012108933879062533
step: 330, loss: 2.6419011192047037e-05
step: 340, loss: 0.0001153927223640494
step: 350, loss: 3.1175994081422687e-05
step: 360, loss: 6.465100159402937e-05
step: 370, loss: 0.0004891653661616147
step: 380, loss: 3.649450809461996e-05
step: 390, loss: 0.00022558419732376933
step: 400, loss: 0.00027674398734234273
step: 410, loss: 0.00010494214075151831
step: 420, loss: 5.254327697912231e-05
step: 430, loss: 5.3005842346465215e-05
step: 440, loss: 3.538528108038008e-05
step: 450, loss: 0.00255717895925045
step: 460, loss: 5.744439840782434e-05
step: 470, loss: 3.093702616752125e-05
step: 480, loss: 0.00014626535994466394
step: 490, loss: 0.03140988573431969
step: 500, loss: 0.00594589626416564
step: 510, loss: 0.0003450941585469991
step: 520, loss: 5.168904681340791e-05
step: 530, loss: 0.00012179102486697957
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9333333333333335, f1=0.9358914365933552,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006947247893549502
step: 10, loss: 0.0012503517791628838
step: 20, loss: 2.4604110876680352e-05
step: 30, loss: 0.00018281240772921592
step: 40, loss: 0.0001614854991203174
step: 50, loss: 0.0008988864719867706
step: 60, loss: 0.0047151632606983185

```

step: 70, loss: 0.0021248036064207554  
step: 80, loss: 0.005381019786000252  
step: 90, loss: 0.0002121470170095563  
step: 100, loss: 0.0016185326967388391  
step: 110, loss: 5.731547571485862e-05  
step: 120, loss: 0.014717387035489082  
step: 130, loss: 2.0037747162859887e-05  
step: 140, loss: 0.0002818132343236357  
step: 150, loss: 0.0009837457910180092  
step: 160, loss: 0.031959012150764465  
step: 170, loss: 8.559012348996475e-05  
step: 180, loss: 2.619914448587224e-05  
step: 190, loss: 4.0915962017606944e-05  
step: 200, loss: 0.009795665740966797  
step: 210, loss: 0.00014100027328822762  
step: 220, loss: 0.00017779337940737605  
step: 230, loss: 4.5456268708221614e-05  
step: 240, loss: 2.7465564926387742e-05  
step: 250, loss: 0.00012084227637387812  
step: 260, loss: 9.1772191808559e-05  
step: 270, loss: 0.00408983463421464  
step: 280, loss: 0.00011342245124978945  
step: 290, loss: 0.0007731548394076526  
step: 300, loss: 0.007155261468142271  
step: 310, loss: 0.07713483273983002  
step: 320, loss: 0.014962990768253803  
step: 330, loss: 0.19936776161193848  
step: 340, loss: 0.00020747519738506526  
step: 350, loss: 0.0022277915850281715  
step: 360, loss: 0.001000096439383924  
step: 370, loss: 3.2478259527124465e-05  
step: 380, loss: 4.2541367292869836e-05  
step: 390, loss: 0.0004793640982825309  
step: 400, loss: 2.4827899324009195e-05  
step: 410, loss: 0.0007291370420716703  
step: 420, loss: 0.00025636644568294287  
step: 430, loss: 0.001365058240480721  
step: 440, loss: 6.730003224220127e-05  
step: 450, loss: 0.018762972205877304  
step: 460, loss: 0.0022470219992101192  
step: 470, loss: 0.0004460396303329617  
step: 480, loss: 7.591842586407438e-05  
step: 490, loss: 0.003104953560978174  
step: 500, loss: 0.00011587091285036877  
step: 510, loss: 5.780237916042097e-05  
step: 520, loss: 2.752476211753674e-05  
step: 530, loss: 3.995028237113729e-05  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9355140186915888, f1=0.9327102803738317,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004667626228183508
step: 10, loss: 0.002143785823136568
step: 20, loss: 2.5498653485556133e-05
step: 30, loss: 4.7726029151817784e-05
step: 40, loss: 0.005279011093080044
step: 50, loss: 0.0027096790727227926
step: 60, loss: 0.03913521021604538
step: 70, loss: 0.00014988119073677808
step: 80, loss: 0.002014301950111985
step: 90, loss: 0.0009294968331232667
step: 100, loss: 7.453641592292115e-05
step: 110, loss: 5.808184505440295e-05
step: 120, loss: 0.001152611686848104
step: 130, loss: 0.0002118974516633898
step: 140, loss: 4.002224522992037e-05
step: 150, loss: 0.0022326877806335688
step: 160, loss: 7.085864490363747e-05
step: 170, loss: 0.0001663511065999046
step: 180, loss: 0.0010970316361635923
step: 190, loss: 0.0014276556903496385
step: 200, loss: 0.00025760504649952054
step: 210, loss: 9.375368972541764e-05
step: 220, loss: 0.0006845702300779521
step: 230, loss: 0.004420312121510506
step: 240, loss: 0.00011737485328922048
step: 250, loss: 5.9946327382931486e-05
step: 260, loss: 0.0006965101347304881
step: 270, loss: 3.238552017137408e-05
step: 280, loss: 4.914806049782783e-05
step: 290, loss: 0.0015813105273991823
step: 300, loss: 7.409232784993947e-05
step: 310, loss: 0.00012449391942936927
step: 320, loss: 0.0002987048646900803
step: 330, loss: 3.33212228724733e-05
step: 340, loss: 6.413515075109899e-05
step: 350, loss: 0.02554386667907238
step: 360, loss: 0.0005009173764847219
step: 370, loss: 5.7789711718214676e-05
step: 380, loss: 5.7849480072036386e-05
step: 390, loss: 2.90895677608205e-05
step: 400, loss: 2.4846982341841795e-05

```

```

step: 410, loss: 0.0008013518527150154
step: 420, loss: 0.0008233332773670554
step: 430, loss: 3.666295378934592e-05
step: 440, loss: 0.00018196880409959704
step: 450, loss: 0.00014901574468240142
step: 460, loss: 0.0001835288421716541
step: 470, loss: 5.556482938118279e-05
step: 480, loss: 0.00014403791283257306
step: 490, loss: 0.00033422556589357555
step: 500, loss: 0.0011264013592153788
step: 510, loss: 4.4603253627428785e-05
step: 520, loss: 0.020436815917491913
step: 530, loss: 0.0003607175895012915
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9354243542435424, f1=0.9383057090239412,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.973718412453309e-05
step: 10, loss: 0.00011556018580449745
step: 20, loss: 0.024000350385904312
step: 30, loss: 0.041466232389211655
step: 40, loss: 0.00012032160157104954
step: 50, loss: 9.533348929835483e-05
step: 60, loss: 3.443883906584233e-05
step: 70, loss: 0.00024080602452158928
step: 80, loss: 3.274013943155296e-05
step: 90, loss: 0.0006908044451847672
step: 100, loss: 2.0477413272601552e-05
step: 110, loss: 6.577458407264203e-05
step: 120, loss: 0.0002992589143104851
step: 130, loss: 0.0020105051808059216
step: 140, loss: 5.5049713409971446e-05
step: 150, loss: 5.530248745344579e-05
step: 160, loss: 9.504971967544407e-05
step: 170, loss: 0.004777992609888315
step: 180, loss: 2.885099638660904e-05
step: 190, loss: 5.587903069681488e-05
step: 200, loss: 0.018725313246250153
step: 210, loss: 0.001104017486795783
step: 220, loss: 2.6824704036698677e-05
step: 230, loss: 1.4841330994386226e-05
step: 240, loss: 0.00032694253604859114
step: 250, loss: 0.034268271178007126
step: 260, loss: 0.0001647678145673126

```

```

step: 270, loss: 2.4399761969107203e-05
step: 280, loss: 6.964516796870157e-05
step: 290, loss: 6.74822658766061e-05
step: 300, loss: 6.368640606524423e-05
step: 310, loss: 0.00015512038953602314
step: 320, loss: 3.0545968911610544e-05
step: 330, loss: 0.0018328094156458974
step: 340, loss: 2.9215971153462306e-05
step: 350, loss: 0.02557945065200329
step: 360, loss: 2.9916171115473844e-05
step: 370, loss: 2.3825117750675417e-05
step: 380, loss: 5.2456172852544114e-05
step: 390, loss: 2.4224977096309885e-05
step: 400, loss: 2.4990844394778833e-05
step: 410, loss: 0.00013132873573340476
step: 420, loss: 2.8768525226041675e-05
step: 430, loss: 0.00022481221822090447
step: 440, loss: 0.000428029743488878
step: 450, loss: 5.544703162740916e-05
step: 460, loss: 2.0231553207850084e-05
step: 470, loss: 0.027725130319595337
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 480, loss: 0.0007151947938837111
step: 490, loss: 0.0002940564590971917
step: 500, loss: 0.00032620731508359313
step: 510, loss: 2.2507696485263295e-05
step: 520, loss: 2.057429992419202e-05
step: 530, loss: 0.0019099077908322215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.936150666054203, f1=0.9382716049382716,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010223829303868115
step: 10, loss: 4.29518440796528e-05
step: 20, loss: 1.6875295841600746e-05
step: 30, loss: 3.647577977972105e-05
step: 40, loss: 4.32067499787081e-05
step: 50, loss: 2.1326752175809816e-05
step: 60, loss: 0.00011986828030785546
step: 70, loss: 4.9116515583591536e-05
step: 80, loss: 0.00046613055747002363
step: 90, loss: 1.533677459519822e-05
step: 100, loss: 4.0011273085838184e-05
step: 110, loss: 3.38986465067137e-05

```

```
step: 120, loss: 4.484973396756686e-05
step: 130, loss: 0.0006423569284379482
step: 140, loss: 0.03826296702027321
step: 150, loss: 5.8962807088391855e-05
step: 160, loss: 0.19626274704933167
step: 170, loss: 0.0010010505793616176
step: 180, loss: 0.047273267060518265
step: 190, loss: 0.0003552331472747028
step: 200, loss: 0.013908088207244873
step: 210, loss: 6.465904880315065e-05
step: 220, loss: 5.974967643851414e-05
step: 230, loss: 0.0012165236985310912
step: 240, loss: 2.9622417059727013e-05
step: 250, loss: 4.822706978302449e-05
step: 260, loss: 3.32622294081375e-05
step: 270, loss: 0.0034050638787448406
step: 280, loss: 2.4600700271548703e-05
step: 290, loss: 6.946456414880231e-05
step: 300, loss: 3.148816540488042e-05
step: 310, loss: 2.943482468253933e-05
step: 320, loss: 0.0001736099657136947
step: 330, loss: 9.151086123893037e-05
step: 340, loss: 3.977280721301213e-05
step: 350, loss: 0.00021520933660212904
step: 360, loss: 0.0022946447134017944
step: 370, loss: 0.00015245788381434977
step: 380, loss: 2.3908027287689038e-05
step: 390, loss: 3.025544538104441e-05
step: 400, loss: 4.591467950376682e-05
step: 410, loss: 3.19907849188894e-05
step: 420, loss: 2.2608097424381413e-05
step: 430, loss: 2.4359051167266443e-05
step: 440, loss: 1.7989046682487242e-05
step: 450, loss: 2.5167335479636677e-05
step: 460, loss: 4.0461207390762866e-05
step: 470, loss: 1.7624039173824713e-05
step: 480, loss: 3.259048753534444e-05
step: 490, loss: 2.211613900726661e-05
step: 500, loss: 0.0004156996728852391
step: 510, loss: 3.20770159305539e-05
step: 520, loss: 5.0672912038862705e-05
step: 530, loss: 0.0007984981057234108
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9354536950420954, f1=0.9350163627863488,
best_f1=0.9358914365933552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 5.6891665735747665e-05
step: 10, loss: 0.00018534171977080405
step: 20, loss: 2.9677819838980213e-05
step: 30, loss: 0.0002423135010758415
step: 40, loss: 3.1992196454666555e-05
step: 50, loss: 0.0017924830317497253
step: 60, loss: 2.0693065380328335e-05
step: 70, loss: 0.0006144656799733639
step: 80, loss: 3.710008968482725e-05
step: 90, loss: 2.8543929147417657e-05
step: 100, loss: 3.2737181754782796e-05
step: 110, loss: 0.009844300337135792
step: 120, loss: 0.0015420063864439726
step: 130, loss: 2.402699101367034e-05
step: 140, loss: 1.5761448594275862e-05
step: 150, loss: 0.00015762497787363827
step: 160, loss: 2.7055775717599317e-05
step: 170, loss: 4.0746475860942155e-05
step: 180, loss: 0.0008000025991350412
step: 190, loss: 5.187121860217303e-05
step: 200, loss: 3.191620635334402e-05
step: 210, loss: 5.009139931644313e-05
step: 220, loss: 2.049582872132305e-05
step: 230, loss: 2.669430796231609e-05
step: 240, loss: 0.0009731273748911917
step: 250, loss: 0.07522072643041611
step: 260, loss: 0.0001060063877957873
step: 270, loss: 0.0002887023729272187
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 280, loss: 0.0019651763141155243
step: 290, loss: 0.001455648336559534
step: 300, loss: 0.02644597738981247
step: 310, loss: 8.537779649486765e-05
step: 320, loss: 0.0005048711900599301
step: 330, loss: 6.831065547885373e-05
step: 340, loss: 0.00017594093515072018
step: 350, loss: 0.0002242708287667483
step: 360, loss: 7.392061525024474e-05
step: 370, loss: 5.3642150305677205e-05
step: 380, loss: 2.323745866306126e-05
step: 390, loss: 2.0942954506608658e-05
step: 400, loss: 2.1166410078876652e-05
step: 410, loss: 0.0010168825974687934
step: 420, loss: 3.424876558710821e-05
step: 430, loss: 2.482083800714463e-05
step: 440, loss: 0.23047220706939697
```

```

step: 450, loss: 0.0031543823424726725
step: 460, loss: 0.00019061386410612613
step: 470, loss: 6.957867299206555e-05
step: 480, loss: 0.0004901720676571131
step: 490, loss: 4.3474065023474395e-05
step: 500, loss: 3.731789183802903e-05
step: 510, loss: 3.937996007152833e-05
step: 520, loss: 2.1796167857246473e-05
step: 530, loss: 4.238163455738686e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.937528921795465, f1=0.9377018920166128,
best_f1=0.9358914365933552
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:21, 266.03it/s]
load_f1 = 0.9390980939098094
real_f1 = 0.9377901578458682
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:16, 268.81it/s]
```

## 2.1.4 Amazon-Google - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Amazon-Google \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8507401943206787
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.0658070296049118
step: 20, loss: 0.3754103481769562
step: 30, loss: 0.37388360500335693
step: 40, loss: 0.48095372319221497
step: 50, loss: 0.2931322157382965
step: 60, loss: 0.3688461184501648
step: 70, loss: 0.2347387969493866
step: 80, loss: 0.3023688495159149
step: 90, loss: 0.36325085163116455
step: 100, loss: 0.12379767745733261
step: 110, loss: 0.2843015789985657
step: 120, loss: 0.19390323758125305
step: 130, loss: 0.22192619740962982
step: 140, loss: 0.21261759102344513
step: 150, loss: 0.233341246843338
step: 160, loss: 0.25879478454589844
step: 170, loss: 0.13235299289226532
step: 180, loss: 0.170585036277771
step: 190, loss: 0.2901453971862793
step: 200, loss: 0.15761475265026093
step: 210, loss: 0.39262935519218445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6037099494097807, f1=0.6292947558770343,
best_f1=0.6292947558770343
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08461623638868332
step: 10, loss: 0.0810711681842804
step: 20, loss: 0.3302020728588104
step: 30, loss: 0.06782863289117813
step: 40, loss: 0.1484336405992508
step: 50, loss: 0.17525480687618256
step: 60, loss: 0.036926738917827606
step: 70, loss: 0.11611762642860413
step: 80, loss: 0.15188893675804138
step: 90, loss: 0.08013227581977844
step: 100, loss: 0.08871544152498245
step: 110, loss: 0.10700438171625137
step: 120, loss: 0.23347438871860504
step: 130, loss: 0.21747122704982758
step: 140, loss: 0.16702425479888916
step: 150, loss: 0.1483151763677597
step: 160, loss: 0.2137105017900467
step: 170, loss: 0.21562768518924713
step: 180, loss: 0.22005122900009155
step: 190, loss: 0.16304348409175873
step: 200, loss: 0.23085127770900726
step: 210, loss: 0.11705267429351807
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6570841889117043, f1=0.6443514644351465,
best_f1=0.6443514644351465
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09247159957885742
step: 10, loss: 0.20656298100948334
step: 20, loss: 0.19853337109088898

```

```

step: 30, loss: 0.0591425783932209
step: 40, loss: 0.15780341625213623
step: 50, loss: 0.16162943840026855
step: 60, loss: 0.20817351341247559
step: 70, loss: 0.11990825831890106
step: 80, loss: 0.05983392521739006
step: 90, loss: 0.18507613241672516
step: 100, loss: 0.02382354624569416
step: 110, loss: 0.08210515975952148
step: 120, loss: 0.21460837125778198
step: 130, loss: 0.07228994369506836
step: 140, loss: 0.2902981638908386
step: 150, loss: 0.26990771293640137
step: 160, loss: 0.10086692124605179
step: 170, loss: 0.10969464480876923
step: 180, loss: 0.1428878903388977
step: 190, loss: 0.14538177847862244
step: 200, loss: 0.11229537427425385
step: 210, loss: 0.15198665857315063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6706827309236948, f1=0.6569037656903767,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08126707375049591
step: 10, loss: 0.013531461358070374
step: 20, loss: 0.17818807065486908
step: 30, loss: 0.03836998715996742
step: 40, loss: 0.08808333426713943
step: 50, loss: 0.04770616814494133
step: 60, loss: 0.11188378185033798
step: 70, loss: 0.17190305888652802
step: 80, loss: 0.07256603240966797
step: 90, loss: 0.03600170463323593
step: 100, loss: 0.04234761744737625
step: 110, loss: 0.08159136027097702
step: 120, loss: 0.018625950440764427
step: 130, loss: 0.10020069032907486
step: 140, loss: 0.17204055190086365
step: 150, loss: 0.14198368787765503
step: 160, loss: 0.013472940772771835

```

```

step: 170, loss: 0.023020144551992416
step: 180, loss: 0.13494296371936798
step: 190, loss: 0.1819969266653061
step: 200, loss: 0.045645151287317276
step: 210, loss: 0.01666233316063881
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.6321626617375232, f1=0.6341463414634146,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05500620976090431
step: 10, loss: 0.049048732966184616
step: 20, loss: 0.017205700278282166
step: 30, loss: 0.14341631531715393
step: 40, loss: 0.10679786652326584
step: 50, loss: 0.08496802300214767
step: 60, loss: 0.0137045131996274
step: 70, loss: 0.284787118434906
step: 80, loss: 0.04775966703891754
step: 90, loss: 0.0055108582600951195
step: 100, loss: 0.0451589860022068
step: 110, loss: 0.0051796515472233295
step: 120, loss: 0.010397466830909252
step: 130, loss: 0.07507681846618652
step: 140, loss: 0.027464304119348526
step: 150, loss: 0.03413378819823265
step: 160, loss: 0.03162321448326111
step: 170, loss: 0.016369903460144997
step: 180, loss: 0.045105792582035065
step: 190, loss: 0.20153537392616272
step: 200, loss: 0.18452803790569305
step: 210, loss: 0.060357462614774704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.6653696498054474, f1=0.6641074856046065,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.052954573184251785
step: 10, loss: 0.01633959449827671
step: 20, loss: 0.0018726594280451536
step: 30, loss: 0.09286994487047195
step: 40, loss: 0.0016051720594987273

```

```

step: 50, loss: 0.0232368353754282
step: 60, loss: 0.15357841551303864
step: 70, loss: 0.002330818912014365
step: 80, loss: 0.11552675068378448
step: 90, loss: 0.04889334365725517
step: 100, loss: 0.03764021024107933
step: 110, loss: 0.01463056355714798
step: 120, loss: 0.010987980291247368
step: 130, loss: 0.011500771157443523
step: 140, loss: 0.041597411036491394
step: 150, loss: 0.009567322209477425
step: 160, loss: 0.1089438796043396
step: 170, loss: 0.006803714204579592
step: 180, loss: 0.005406027659773827
step: 190, loss: 0.006491277366876602
step: 200, loss: 0.09646143019199371
step: 210, loss: 0.06736980378627777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6691042047531993, f1=0.6607142857142858,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08570700138807297
step: 10, loss: 0.012895406223833561
step: 20, loss: 0.026624297723174095
step: 30, loss: 0.054178331047296524
step: 40, loss: 0.027209332212805748
step: 50, loss: 0.06269554793834686
step: 60, loss: 0.10815764963626862
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 70, loss: 0.0042730458080768585
step: 80, loss: 0.024185050278902054
step: 90, loss: 0.0016407400835305452
step: 100, loss: 0.004525292199105024
step: 110, loss: 0.07364889979362488
step: 120, loss: 0.08526723831892014
step: 130, loss: 0.0037616596091538668
step: 140, loss: 0.0417628288269043
step: 150, loss: 0.010626141913235188
step: 160, loss: 0.019463540986180305
step: 170, loss: 0.017406797036528587
step: 180, loss: 0.006373814307153225
step: 190, loss: 0.034407224506139755
step: 200, loss: 0.07356144487857819
step: 210, loss: 0.03276824951171875

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.6583184257602862, f1=0.6433566433566433,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01557499822229147
step: 10, loss: 0.0011982866562902927
step: 20, loss: 0.0016105377580970526
step: 30, loss: 0.005833758041262627
step: 40, loss: 0.041766900569200516
step: 50, loss: 0.0029038856737315655
step: 60, loss: 0.016889583319425583
step: 70, loss: 0.0022409954108297825
step: 80, loss: 0.011874333024024963
step: 90, loss: 0.026331964880228043
step: 100, loss: 0.0006027221097610891
step: 110, loss: 0.002312560798600316
step: 120, loss: 0.08748750388622284
step: 130, loss: 0.003623973112553358
step: 140, loss: 0.021624376997351646
step: 150, loss: 0.018928179517388344
step: 160, loss: 0.08216299116611481
step: 170, loss: 0.006356530822813511
step: 180, loss: 0.027432756498456
step: 190, loss: 0.0591595396399498
step: 200, loss: 0.031455814838409424
step: 210, loss: 0.05567427724599838
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.6561886051080551, f1=0.6475095785440614,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006538189947605133
step: 10, loss: 0.004396774340420961
step: 20, loss: 0.06774502247571945
step: 30, loss: 0.03467060625553131
step: 40, loss: 0.01363541278988123
step: 50, loss: 0.0025586115662008524
step: 60, loss: 0.04459325224161148
step: 70, loss: 0.005295020528137684
step: 80, loss: 0.09978825598955154
step: 90, loss: 0.06175601854920387

```

```

step: 100, loss: 0.03616424277424812
step: 110, loss: 0.0012190293055027723
step: 120, loss: 0.01888353005051613
step: 130, loss: 0.03606272488832474
step: 140, loss: 0.013322245329618454
step: 150, loss: 0.001356263179332018
step: 160, loss: 0.0014853713801130652
step: 170, loss: 0.015785764902830124
step: 180, loss: 0.03186913952231407
step: 190, loss: 0.016609642654657364
step: 200, loss: 0.026768142357468605
step: 210, loss: 0.045094557106494904
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6415094339622642, f1=0.6187845303867404,
best_f1=0.6569037656903767
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005872430745512247
step: 10, loss: 0.00192354922182858
step: 20, loss: 0.026396464556455612
step: 30, loss: 0.0022700480185449123
step: 40, loss: 0.017195412889122963
step: 50, loss: 0.0029941832181066275
step: 60, loss: 0.0005391928716562688
step: 70, loss: 0.026455486193299294
step: 80, loss: 0.0001501387159805745
step: 90, loss: 0.02191263623535633
step: 100, loss: 0.001654390012845397
step: 110, loss: 0.001256540883332491
step: 120, loss: 0.042811207473278046
step: 130, loss: 0.01592613384127617
step: 140, loss: 0.0002390385343460366
step: 150, loss: 0.2651784420013428
step: 160, loss: 0.005352059379220009
step: 170, loss: 0.015710733830928802
step: 180, loss: 0.0010514206951484084
step: 190, loss: 0.06981217116117477
step: 200, loss: 0.016996394842863083
step: 210, loss: 0.0008790276478976011
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.6716141001855287, f1=0.6366906474820143,
best_f1=0.6366906474820143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00861451867967844
step: 10, loss: 0.02979176864027977
step: 20, loss: 0.0005527614848688245
step: 30, loss: 0.00328380661085248
step: 40, loss: 0.015679681673645973
step: 50, loss: 0.0032253889366984367
step: 60, loss: 0.002513529034331441
step: 70, loss: 0.0011167566990479827
step: 80, loss: 0.0758013054728508
step: 90, loss: 0.0008828100399114192
step: 100, loss: 0.0008815007749944925
step: 110, loss: 0.035718630999326706
step: 120, loss: 0.00038106724969111383
step: 130, loss: 0.07305432856082916
step: 140, loss: 0.016825586557388306
step: 150, loss: 0.021168140694499016
step: 160, loss: 0.16503258049488068
step: 170, loss: 0.15110793709754944
step: 180, loss: 0.015349901281297207
step: 190, loss: 0.0009088648366741836
step: 200, loss: 0.0002011681644944474
step: 210, loss: 0.0005628173821605742
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6447876447876447, f1=0.6386233269598471,
best_f1=0.6366906474820143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003320178366266191
step: 10, loss: 0.0022921499330550432
step: 20, loss: 0.0013261731946840882
step: 30, loss: 0.0013210558099672198
step: 40, loss: 0.006568788550794125
step: 50, loss: 0.053496718406677246
step: 60, loss: 0.00037223182152956724
step: 70, loss: 0.0018030342180281878
step: 80, loss: 0.0031795529648661613
step: 90, loss: 0.005212148651480675
step: 100, loss: 0.00034824732574634254
step: 110, loss: 0.00040345650631934404

```

```

step: 120, loss: 0.0006343983695842326
step: 130, loss: 0.010942762717604637
step: 140, loss: 0.0006093030679039657
step: 150, loss: 0.0003160069463774562
step: 160, loss: 0.003626641584560275
step: 170, loss: 0.010549853555858135
step: 180, loss: 0.00030379282543435693
step: 190, loss: 0.016780512407422066
step: 200, loss: 0.026458600535988808
step: 210, loss: 0.0041280388832092285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6509803921568627, f1=0.6472868217054264,
best_f1=0.6366906474820143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007592943729832768
step: 10, loss: 0.0022316414397209883
step: 20, loss: 0.00044560522655956447
step: 30, loss: 0.03867137059569359
step: 40, loss: 0.0002156569971702993
step: 50, loss: 0.0007216008380055428
step: 60, loss: 0.0003234791220165789
step: 70, loss: 0.007376214489340782
step: 80, loss: 0.016526155173778534
step: 90, loss: 0.038051459938287735
step: 100, loss: 0.0648861974477768
step: 110, loss: 0.00018712846213020384
step: 120, loss: 0.0007088775746524334
step: 130, loss: 0.00020471516472753137
step: 140, loss: 0.0017109573818743229
step: 150, loss: 0.0004146759456489235
step: 160, loss: 0.04174193739891052
step: 170, loss: 0.0007310193032026291
step: 180, loss: 0.034096863120794296
step: 190, loss: 0.0009920246666297317
step: 200, loss: 0.0006737521034665406
step: 210, loss: 0.015431905165314674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6467065868263474, f1=0.6614481409001958,
best_f1=0.6366906474820143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.0020628543570637703
step: 10, loss: 0.0001949126599356532
step: 20, loss: 0.00010419724276289344
step: 30, loss: 0.001932394108735025
step: 40, loss: 0.0035165443550795317
step: 50, loss: 0.013747338205575943
step: 60, loss: 0.0005147674819454551
step: 70, loss: 0.10502036660909653
step: 80, loss: 0.030623644590377808
step: 90, loss: 0.06236843764781952
step: 100, loss: 0.0017856562044471502
step: 110, loss: 0.0012767839943990111
step: 120, loss: 0.0101866340264678
step: 130, loss: 0.0004947113338857889
step: 140, loss: 0.0001550318265799433
step: 150, loss: 0.0002906535810325295
step: 160, loss: 0.0003202286025043577
step: 170, loss: 0.001332951826043427
step: 180, loss: 0.011138113215565681
step: 190, loss: 0.0001717878767522052
step: 200, loss: 0.0004045585810672492
step: 210, loss: 0.00025950450799427927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6524271844660194, f1=0.6449136276391555,
best_f1=0.6366906474820143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014651167206466198
step: 10, loss: 0.00791336689144373
step: 20, loss: 0.000183919066330418
step: 30, loss: 0.0002777192275971174
step: 40, loss: 0.0008243990014307201
step: 50, loss: 0.00013602299441117793
step: 60, loss: 0.0004301567387301475
step: 70, loss: 0.00042088699410669506
step: 80, loss: 0.00016735185636207461
step: 90, loss: 0.0019497537286952138
step: 100, loss: 0.00041106101707555354
step: 110, loss: 0.00013433773710858077
step: 120, loss: 0.00029818963957950473
step: 130, loss: 0.003094144631177187
step: 140, loss: 0.000320965627906844
step: 150, loss: 0.0002932412317022681
step: 160, loss: 0.0061947256326675415
step: 170, loss: 0.036017052829265594

```

```

step: 180, loss: 0.0009169570403173566
step: 190, loss: 0.009087606333196163
step: 200, loss: 0.019441040232777596
step: 210, loss: 0.0011750562116503716
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6365503080082136, f1=0.6434426229508197,
best_f1=0.6366906474820143
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:06, 350.86it/s]
load_f1 = 0.674074074074074
real_f1 = 0.6703703703703704
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:16, 265.62it/s]

```

### 2.1.5 Walmart-Amazon - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Walmart-Amazon \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/Walmart-Amazon \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: O2
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

enabled	: True
opt_level	: O2
cast_model_type	: torch.float16
patch_torch_functions	: False

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8591319918632507
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1719321459531784
step: 20, loss: 0.1527455449104309
step: 30, loss: 0.5192669630050659
step: 40, loss: 0.2701660990715027
step: 50, loss: 0.3096366226673126
step: 60, loss: 0.3599570393562317
step: 70, loss: 0.17629890143871307
step: 80, loss: 0.5198954939842224
step: 90, loss: 0.23786242306232452
step: 100, loss: 0.21916985511779785
step: 110, loss: 0.23342619836330414
step: 120, loss: 0.42488110065460205
step: 130, loss: 0.3407513201236725
step: 140, loss: 0.3270037770271301
step: 150, loss: 0.2720862627029419
step: 160, loss: 0.21450330317020416
step: 170, loss: 0.40535807609558105
step: 180, loss: 0.2897043228149414
step: 190, loss: 0.13775089383125305
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.39627039627039634, f1=0.4243792325056433,

```

```

best_f1=0.4243792325056433
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.30331170558929443
step: 10, loss: 0.08539453893899918
step: 20, loss: 0.1864873617887497
step: 30, loss: 0.24919666349887848
step: 40, loss: 0.5005083084106445
step: 50, loss: 0.2712303102016449
step: 60, loss: 0.19003444910049438
step: 70, loss: 0.1689527928829193
step: 80, loss: 0.1404629945755005
step: 90, loss: 0.09818045794963837
step: 100, loss: 0.3000776469707489
step: 110, loss: 0.13529109954833984
step: 120, loss: 0.37140092253685
step: 130, loss: 0.11695404350757599
step: 140, loss: 0.22801630198955536
step: 150, loss: 0.02836550958454609
step: 160, loss: 0.11246450245380402
step: 170, loss: 0.29932913184165955
step: 180, loss: 0.2031008005142212
step: 190, loss: 0.13929694890975952
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6900269541778975, f1=0.7142857142857142,
best_f1=0.7142857142857142
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15526720881462097
step: 10, loss: 0.19634398818016052
step: 20, loss: 0.01423700526356697
step: 30, loss: 0.02316397614777088
step: 40, loss: 0.10939715802669525
step: 50, loss: 0.051510054618120193
step: 60, loss: 0.08493571728467941
step: 70, loss: 0.10761206597089767
step: 80, loss: 0.23152194917201996
step: 90, loss: 0.3424510359764099
step: 100, loss: 0.17351743578910828
step: 110, loss: 0.14443856477737427

```

```

step: 120, loss: 0.09400435537099838
step: 130, loss: 0.1491655707359314
step: 140, loss: 0.04817602410912514
step: 150, loss: 0.07945292443037033
step: 160, loss: 0.05933864787220955
step: 170, loss: 0.0630510002374649
step: 180, loss: 0.03853816166520119
step: 190, loss: 0.21100832521915436
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7344632768361581, f1=0.7138643067846608,
best_f1=0.7138643067846608
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012059430591762066
step: 10, loss: 0.01762687973678112
step: 20, loss: 0.012513731606304646
step: 30, loss: 0.045875199139118195
step: 40, loss: 0.09516175091266632
step: 50, loss: 0.12135497480630875
step: 60, loss: 0.052459172904491425
step: 70, loss: 0.005354410037398338
step: 80, loss: 0.10521107167005539
step: 90, loss: 0.23010367155075073
step: 100, loss: 0.015247631818056107
step: 110, loss: 0.01181950606405735
step: 120, loss: 0.09526726603507996
step: 130, loss: 0.2856878638267517
step: 140, loss: 0.03133070096373558
step: 150, loss: 0.06991373747587204
step: 160, loss: 0.10849232971668243
step: 170, loss: 0.009120098315179348
step: 180, loss: 0.08113770931959152
step: 190, loss: 0.03843569755554199
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7552083333333333, f1=0.7393617021276595,

```

```

best_f1=0.7393617021276595
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04367104917764664
step: 10, loss: 0.021557295694947243
step: 20, loss: 0.02552308887243271
step: 30, loss: 0.003500333521515131
step: 40, loss: 0.009220534935593605
step: 50, loss: 0.022138191387057304
step: 60, loss: 0.026563169434666634
step: 70, loss: 0.0006669930298812687
step: 80, loss: 0.023057855665683746
step: 90, loss: 0.008126366883516312
step: 100, loss: 0.03095051646232605
step: 110, loss: 0.03406832739710808
step: 120, loss: 0.0006631079595535994
step: 130, loss: 0.009687172248959541
step: 140, loss: 0.005736729595810175
step: 150, loss: 0.012456223368644714
step: 160, loss: 0.010808691382408142
step: 170, loss: 0.03530406951904297
step: 180, loss: 0.05514267086982727
step: 190, loss: 0.09838893264532089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7639257294429709, f1=0.7342465753424658,
best_f1=0.7342465753424658
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056511349976062775
step: 10, loss: 0.0010047146352007985
step: 20, loss: 0.0008532613865099847
step: 30, loss: 0.005230030510574579
step: 40, loss: 0.004198548384010792
step: 50, loss: 0.05767573043704033
step: 60, loss: 0.00022727578470949084
step: 70, loss: 0.017810996621847153
step: 80, loss: 0.28148704767227173
step: 90, loss: 0.08163674175739288
step: 100, loss: 0.002466336591169238
step: 110, loss: 0.08893417567014694

```

```

step: 120, loss: 0.07176471501588821
step: 130, loss: 0.022658497095108032
step: 140, loss: 0.003362033050507307
step: 150, loss: 0.047872982919216156
step: 160, loss: 0.001928709913045168
step: 170, loss: 0.02118956483900547
step: 180, loss: 0.015393776819109917
step: 190, loss: 0.10129343718290329
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.7793696275071633, f1=0.750733137829912,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00275000580586493
step: 10, loss: 0.004293602891266346
step: 20, loss: 0.0013005256187170744
step: 30, loss: 0.007647744845598936
step: 40, loss: 0.0023396043106913567
step: 50, loss: 0.2346428632736206
step: 60, loss: 0.005782575346529484
step: 70, loss: 0.0009955177083611488
step: 80, loss: 0.008470458909869194
step: 90, loss: 0.0014192955568432808
step: 100, loss: 0.004770458210259676
step: 110, loss: 0.010073180310428143
step: 120, loss: 0.017833054065704346
step: 130, loss: 0.004678584169596434
step: 140, loss: 0.00312646571546793
step: 150, loss: 0.03760826960206032
step: 160, loss: 0.002691148780286312
step: 170, loss: 0.0011543791042640805
step: 180, loss: 0.015536059625446796
step: 190, loss: 0.042456384748220444
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7444444444444444, f1=0.7409470752089136,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.0008115393575280905
step: 10, loss: 0.01168848667293787
step: 20, loss: 0.0014030245365574956
step: 30, loss: 0.006462109740823507
step: 40, loss: 0.00187685654964298
step: 50, loss: 0.0010058159241452813
step: 60, loss: 0.002054635202512145
step: 70, loss: 0.008768157102167606
step: 80, loss: 0.05657287687063217
step: 90, loss: 0.0012223748490214348
step: 100, loss: 0.0031277271918952465
step: 110, loss: 0.002089966554194689
step: 120, loss: 0.00038034978206269443
step: 130, loss: 0.07462599873542786
step: 140, loss: 0.001949250465258956
step: 150, loss: 0.015801820904016495
step: 160, loss: 0.013269132003188133
step: 170, loss: 0.0005662949988618493
step: 180, loss: 0.0006699266959913075
step: 190, loss: 0.002884515793994069
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7327823691460055, f1=0.7062146892655367,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00032374943839386106
step: 10, loss: 0.0016515752067789435
step: 20, loss: 0.04351702332496643
step: 30, loss: 0.0018949320074170828
step: 40, loss: 0.0035225837491452694
step: 50, loss: 0.002367722801864147
step: 60, loss: 0.000531203462742269
step: 70, loss: 0.003383014816790819
step: 80, loss: 0.005084735341370106
step: 90, loss: 0.022203391417860985
step: 100, loss: 0.0009098952286876738
step: 110, loss: 0.0003367939207237214
step: 120, loss: 0.08252902328968048
step: 130, loss: 0.009904619306325912
step: 140, loss: 0.0001537023636046797
step: 150, loss: 0.0985589250922203
step: 160, loss: 0.0006130951805971563
step: 170, loss: 0.001987996743991971
step: 180, loss: 0.0016219194512814283
step: 190, loss: 0.0027335318736732006

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7262247838616714, f1=0.7058823529411764,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003757033497095108
step: 10, loss: 0.010359568521380424
step: 20, loss: 0.00028890499379485846
step: 30, loss: 0.0015407463070005178
step: 40, loss: 0.0010791599052026868
step: 50, loss: 0.0006895210826769471
step: 60, loss: 0.0046812077052891254
step: 70, loss: 0.0006402154103852808
step: 80, loss: 0.00026056214119307697
step: 90, loss: 0.0005221490864641964
step: 100, loss: 0.0057219830341637135
step: 110, loss: 0.0004959299694746733
step: 120, loss: 0.06561864912509918
step: 130, loss: 0.0003152690187562257
step: 140, loss: 0.0014532277127727866
step: 150, loss: 0.0003744301793631166
step: 160, loss: 0.00014425315021071583
step: 170, loss: 0.0010107951238751411
step: 180, loss: 0.0003307909064460546
step: 190, loss: 0.0012945955386385322
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7228260869565217, f1=0.724233983286908,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03341969475150108
step: 10, loss: 0.00041067757410928607
step: 20, loss: 0.00031141864019446075
step: 30, loss: 0.0036988991778343916
step: 40, loss: 0.00015884482127148658
step: 50, loss: 0.09589573740959167
step: 60, loss: 0.00026016042102128267
step: 70, loss: 0.00010053025471279398
step: 80, loss: 0.0003277624200563878
step: 90, loss: 0.008032799698412418
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 0.0030005015432834625

```

```

step: 110, loss: 0.042288072407245636
step: 120, loss: 0.0005803698441013694
step: 130, loss: 0.00015226282994262874
step: 140, loss: 0.0024228596594184637
step: 150, loss: 0.00018609745893627405
step: 160, loss: 0.0010284970048815012
step: 170, loss: 0.0004845672519877553
step: 180, loss: 0.0006023278110660613
step: 190, loss: 0.0001621234987396747
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7272727272727272, f1=0.7354497354497355,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012113755656173453
step: 10, loss: 0.0002225119824288413
step: 20, loss: 0.00040867755888029933
step: 30, loss: 0.0010198383824899793
step: 40, loss: 0.0008361024083569646
step: 50, loss: 0.0003179093182552606
step: 60, loss: 0.001084611751139164
step: 70, loss: 0.00010572298197075725
step: 80, loss: 0.00024552977993153036
step: 90, loss: 0.0003575004229787737
step: 100, loss: 0.0006957226432859898
step: 110, loss: 0.0008817157940939069
step: 120, loss: 0.00024288934946525842
step: 130, loss: 9.268941357731819e-05
step: 140, loss: 0.0005968844052404165
step: 150, loss: 0.00016347116616088897
step: 160, loss: 0.048634957522153854
step: 170, loss: 0.0028219979722052813
step: 180, loss: 0.0003325024445075542
step: 190, loss: 0.002924615051597357
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7401574803149608, f1=0.7248677248677249,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00044321795576252043
step: 10, loss: 0.0004896469763480127
step: 20, loss: 0.000466795259853825

```

```

step: 30, loss: 0.0008276799344457686
step: 40, loss: 7.634667417732999e-05
step: 50, loss: 0.00019374287512619048
step: 60, loss: 0.0001281119475606829
step: 70, loss: 0.00026337162125855684
step: 80, loss: 0.0002313491131644696
step: 90, loss: 0.0002111214562319219
step: 100, loss: 0.0002710361732169986
step: 110, loss: 0.0001932633458636701
step: 120, loss: 0.0007313991081900895
step: 130, loss: 0.0002618716680444777
step: 140, loss: 0.0001515302137704566
step: 150, loss: 5.535982199944556e-05
step: 160, loss: 0.0001427164243068546
step: 170, loss: 7.745506445644423e-05
step: 180, loss: 0.0021922506857663393
step: 190, loss: 0.0008619912550784647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7371273712737126, f1=0.7298050139275766,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002893785713240504
step: 10, loss: 0.001051287166774273
step: 20, loss: 0.0683164894580841
step: 30, loss: 0.0007241840939968824
step: 40, loss: 0.0009162917267531157
step: 50, loss: 0.00011165750765940174
step: 60, loss: 0.00023490263265557587
step: 70, loss: 0.00034767077886499465
step: 80, loss: 5.932995554758236e-05
step: 90, loss: 0.00019378590513952076
step: 100, loss: 0.00015443975280504674
step: 110, loss: 0.00032637492404319346
step: 120, loss: 0.0004097337950952351
step: 130, loss: 0.00015138874005060643
step: 140, loss: 0.00032430567080155015
step: 150, loss: 0.00021380058024078608
step: 160, loss: 0.00022583280224353075
step: 170, loss: 0.0007531586452387273
step: 180, loss: 0.0008329269476234913
step: 190, loss: 7.261715654749423e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

epoch 14: dev_f1=0.7458563535911602, f1=0.7231638418079097,
best_f1=0.750733137829912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001006075763143599
step: 10, loss: 8.161042933352292e-05
step: 20, loss: 9.188390686176717e-05
step: 30, loss: 0.12741915881633759
step: 40, loss: 0.00033397754305042326
step: 50, loss: 0.017354590818285942
step: 60, loss: 0.00011227672075619921
step: 70, loss: 0.0003274274931754917
step: 80, loss: 8.623937901575118e-05
step: 90, loss: 0.00040129502303898335
step: 100, loss: 0.0007042891811579466
step: 110, loss: 0.00014149512571748346
step: 120, loss: 0.0015741863753646612
step: 130, loss: 0.0017860562074929476
step: 140, loss: 0.000649890280328691
step: 150, loss: 0.00016856711590662599
step: 160, loss: 0.0011954170186072588
step: 170, loss: 0.0005623336182907224
step: 180, loss: 0.011658787727355957
step: 190, loss: 0.0010813219705596566
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7377049180327869, f1=0.7127071823204421,
best_f1=0.750733137829912
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:08, 231.81it/s]
load_f1 = 0.7736389684813755
real_f1 = 0.7675070028011204
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:16, 267.37it/s]

```

## 2.2 BASELINE TEXTUAL

### 2.2.1 Abt-Buy - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Textual/Abt-Buy \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Textual/Abt-Buy \
    --input_path input/input_small.jsonl \

```

```

--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

```

```
step: 0, loss: 0.8533097505569458
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.2227528989315033
```

```
step: 20, loss: 0.16046379506587982
```

```
step: 30, loss: 0.23496641218662262
```

```

step: 40, loss: 0.3061247169971466
step: 50, loss: 0.3725392818450928
step: 60, loss: 0.4442770481109619
step: 70, loss: 0.30968812108039856
step: 80, loss: 0.25331488251686096
step: 90, loss: 0.3893571197986603
step: 100, loss: 0.2271021008491516
step: 110, loss: 0.18637609481811523
step: 120, loss: 0.5565589666366577
step: 130, loss: 0.40610459446907043
step: 140, loss: 0.47056788206100464
step: 150, loss: 0.1197020635008812
step: 160, loss: 0.34333014488220215
step: 170, loss: 0.24178168177604675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.33199999999999996, f1=0.2916666666666667,
best_f1=0.2916666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.4199289381504059
step: 10, loss: 0.170386403799057
step: 20, loss: 0.30800512433052063
step: 30, loss: 0.2380589097738266
step: 40, loss: 0.14613237977027893
step: 50, loss: 0.17466682195663452
step: 60, loss: 0.07353284955024719
step: 70, loss: 0.2336156964302063
step: 80, loss: 0.15290199220180511
step: 90, loss: 0.16126281023025513
step: 100, loss: 0.16349628567695618
step: 110, loss: 0.24604837596416473
step: 120, loss: 0.034274082630872726
step: 130, loss: 0.0940004512667656
step: 140, loss: 0.15803413093090057
step: 150, loss: 0.12340065091848373
step: 160, loss: 0.12072974443435669
step: 170, loss: 0.04907487705349922
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```



UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.76056338028169, f1=0.779816513761468,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.054785698652267456
step: 10, loss: 0.041662443429231644
step: 20, loss: 0.11842634528875351
step: 30, loss: 0.188019260764122
step: 40, loss: 0.0104496655985713
step: 50, loss: 0.11347655951976776
step: 60, loss: 0.25556376576423645
step: 70, loss: 0.05909918621182442
step: 80, loss: 0.21082797646522522
step: 90, loss: 0.12667137384414673
step: 100, loss: 0.03146843612194061
step: 110, loss: 0.13001906871795654
step: 120, loss: 0.00821442436426878
step: 130, loss: 0.07332096993923187
step: 140, loss: 0.0034092015121132135
step: 150, loss: 0.03515207767486572
step: 160, loss: 0.01875258982181549
step: 170, loss: 0.12783078849315643
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.7518427518427517, f1=0.8057553956834533,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.057483967393636703
step: 10, loss: 0.05325731262564659
step: 20, loss: 0.011589628644287586
step: 30, loss: 0.02816314809024334
step: 40, loss: 0.007246298249810934
step: 50, loss: 0.021303189918398857
step: 60, loss: 0.03468302637338638
step: 70, loss: 0.031664229929447174
step: 80, loss: 0.011869811452925205
step: 90, loss: 0.20712913572788239
step: 100, loss: 0.014036537148058414
step: 110, loss: 0.020129499956965446
step: 120, loss: 0.13310272991657257
step: 130, loss: 0.019808480516076088
```

```

step: 140, loss: 0.006757833994925022
step: 150, loss: 0.003794064512476325
step: 160, loss: 0.03777889162302017
step: 170, loss: 0.02739342488348484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7677261613691931, f1=0.8086124401913874,
best_f1=0.8086124401913874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030497882515192032
step: 10, loss: 0.04339165240526199
step: 20, loss: 0.004562702029943466
step: 30, loss: 0.05891839787364006
step: 40, loss: 0.021093107759952545
step: 50, loss: 0.0047235433012247086
step: 60, loss: 0.00264483829960227
step: 70, loss: 0.017774851992726326
step: 80, loss: 0.0039048148319125175
step: 90, loss: 0.030268242582678795
step: 100, loss: 0.035064321011304855
step: 110, loss: 0.24032974243164062
step: 120, loss: 0.07201789319515228
step: 130, loss: 0.005468454677611589
step: 140, loss: 0.02834540791809559
step: 150, loss: 0.014486361294984818
step: 160, loss: 0.013513204641640186
step: 170, loss: 0.04472930729389191
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7543424317617866, f1=0.7848699763593381,
best_f1=0.8086124401913874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00816493108868599
step: 10, loss: 0.04292043671011925
step: 20, loss: 0.007316218689084053
step: 30, loss: 0.07791576534509659
step: 40, loss: 0.0027619758620858192
step: 50, loss: 0.01915605179965496

```

```

step: 60, loss: 0.03680424019694328
step: 70, loss: 0.061817243695259094
step: 80, loss: 0.008325789123773575
step: 90, loss: 0.008735661394894123
step: 100, loss: 0.03553183004260063
step: 110, loss: 0.029950354248285294
step: 120, loss: 0.0047910274006426334
step: 130, loss: 0.10764361172914505
step: 140, loss: 0.0037279354874044657
step: 150, loss: 0.16269370913505554
step: 160, loss: 0.02259671315550804
step: 170, loss: 0.05965277552604675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.7729468599033817, f1=0.8045977011494253,
best_f1=0.8045977011494253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017226967960596085
step: 10, loss: 0.0009849430061876774
step: 20, loss: 0.013073726557195187
step: 30, loss: 0.016217423602938652
step: 40, loss: 0.01300230622291565
step: 50, loss: 0.0005965409800410271
step: 60, loss: 0.17291675508022308
step: 70, loss: 0.0019576866179704666
step: 80, loss: 0.013705745339393616
step: 90, loss: 0.0030870845075696707
step: 100, loss: 0.04722702503204346
step: 110, loss: 0.0003774987126234919
step: 120, loss: 0.04526110738515854
step: 130, loss: 0.131617933511734
step: 140, loss: 0.004170864820480347
step: 150, loss: 0.03395386040210724
step: 160, loss: 0.019582102075219154
step: 170, loss: 0.22933143377304077
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7559808612440191, f1=0.7863636363636364,
best_f1=0.8045977011494253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02054751291871071
step: 10, loss: 0.0018427434843033552
step: 20, loss: 0.019939562305808067
step: 30, loss: 0.054216451942920685
step: 40, loss: 0.0028514680452644825
step: 50, loss: 0.0014813238522037864
step: 60, loss: 0.007617009337991476
step: 70, loss: 0.01834007538855076
step: 80, loss: 0.002336119534447789
step: 90, loss: 0.005405094940215349
step: 100, loss: 0.011577585712075233
step: 110, loss: 0.002525458810850978
step: 120, loss: 0.00299801304936409
step: 130, loss: 0.00190757738891989
step: 140, loss: 0.0018290594452992082
step: 150, loss: 0.09059051424264908
step: 160, loss: 0.021474990993738174
step: 170, loss: 0.12259584665298462
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.7811764705882352, f1=0.7963386727688786,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032721564639359713
step: 10, loss: 0.0012254740577191114
step: 20, loss: 0.002662333194166422
step: 30, loss: 0.07894903421401978
step: 40, loss: 0.015221059322357178
step: 50, loss: 0.03826924040913582
step: 60, loss: 0.0006523799384012818
step: 70, loss: 0.0031835180707275867
step: 80, loss: 0.05009506270289421
step: 90, loss: 0.0013701488496735692
step: 100, loss: 0.02060690149664879
step: 110, loss: 0.0003955152933485806
step: 120, loss: 0.012394100427627563
step: 130, loss: 0.0013556863414123654
step: 140, loss: 0.04326607659459114
step: 150, loss: 0.01798636093735695

```

```

step: 160, loss: 0.007364185526967049
step: 170, loss: 0.007140072528272867
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7737226277372263, f1=0.7962962962962963,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008404645137488842
step: 10, loss: 0.003629636485129595
step: 20, loss: 0.0004862983478233218
step: 30, loss: 0.0002785962133202702
step: 40, loss: 0.001920598791912198
step: 50, loss: 0.0014044326962903142
step: 60, loss: 0.0015189830446615815
step: 70, loss: 0.09516146779060364
step: 80, loss: 0.020986465737223625
step: 90, loss: 0.032377831637859344
step: 100, loss: 0.0041250307112932205
step: 110, loss: 0.002334660617634654
step: 120, loss: 0.018520524725317955
step: 130, loss: 0.03322058916091919
step: 140, loss: 0.05890890583395958
step: 150, loss: 0.000751435523852706
step: 160, loss: 0.000222629846073687
step: 170, loss: 0.005201131105422974
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7738927738927739, f1=0.7675438596491228,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003149354597553611
step: 10, loss: 0.013811543583869934
step: 20, loss: 0.08052314817905426
step: 30, loss: 0.0006168907275423408
step: 40, loss: 0.006545602809637785
step: 50, loss: 0.03237174078822136
step: 60, loss: 0.00460835313424468
step: 70, loss: 0.00317613547667861
step: 80, loss: 0.0013024461222812533
step: 90, loss: 0.002644467167556286
step: 100, loss: 0.0002933118666987866
step: 110, loss: 0.0043062251061201096

```

```

step: 120, loss: 0.0017876868369057775
step: 130, loss: 0.009569655172526836
step: 140, loss: 0.018300814554095268
step: 150, loss: 0.002004428068175912
step: 160, loss: 0.0003983526839874685
step: 170, loss: 0.047986146062612534
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7769784172661871, f1=0.7945205479452053,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04099494591355324
step: 10, loss: 0.005130769684910774
step: 20, loss: 0.03865811228752136
step: 30, loss: 0.00151268660556525
step: 40, loss: 0.00039501837454736233
step: 50, loss: 0.0027573578990995884
step: 60, loss: 0.0006510998937301338
step: 70, loss: 0.004468258935958147
step: 80, loss: 0.10635533928871155
step: 90, loss: 0.0009992808336392045
step: 100, loss: 0.0019002349581569433
step: 110, loss: 0.00020323980425018817
step: 120, loss: 0.07884538173675537
step: 130, loss: 0.00020691764075309038
step: 140, loss: 0.003811844391748309
step: 150, loss: 0.0028206759598106146
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.02264450676739216
step: 170, loss: 0.017555736005306244
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7584541062801932, f1=0.7925407925407925,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008608286269009113
step: 10, loss: 0.05407208949327469
step: 20, loss: 0.0791458785533905
step: 30, loss: 0.00559763191267848
step: 40, loss: 0.0002900154213421047
step: 50, loss: 0.005241654347628355
step: 60, loss: 0.0004035470774397254

```

```

step: 70, loss: 0.01072138361632824
step: 80, loss: 0.0017361192731186748
step: 90, loss: 0.0002255880244774744
step: 100, loss: 0.00025578885106369853
step: 110, loss: 0.0003953921259380877
step: 120, loss: 0.00022691524645779282
step: 130, loss: 0.0018410346237942576
step: 140, loss: 0.0008724177605472505
step: 150, loss: 0.0003092737460974604
step: 160, loss: 0.0014059508685022593
step: 170, loss: 0.006202321499586105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7642679900744417, f1=0.7990654205607476,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002258412423543632
step: 10, loss: 0.000393589842133224
step: 20, loss: 0.005630738101899624
step: 30, loss: 0.017544912174344063
step: 40, loss: 0.00046505386126227677
step: 50, loss: 0.0003332349006086588
step: 60, loss: 0.00016962781955953687
step: 70, loss: 0.0002978518314193934
step: 80, loss: 0.0009409910417161882
step: 90, loss: 0.00018734665354713798
step: 100, loss: 0.00045684666838496923
step: 110, loss: 0.0018990333192050457
step: 120, loss: 0.0048730638809502125
step: 130, loss: 0.005903115030378103
step: 140, loss: 0.0023420555517077446
step: 150, loss: 0.019026121124625206
step: 160, loss: 0.029538443312048912
step: 170, loss: 0.0007294603856280446
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7775175644028103, f1=0.7918552036199096,
best_f1=0.7963386727688786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00034504488576203585
step: 10, loss: 0.004500280134379864
step: 20, loss: 0.0017965100705623627

```

```

step: 30, loss: 0.034895751625299454
step: 40, loss: 0.0001705557224340737
step: 50, loss: 0.00011170420475536957
step: 60, loss: 0.00019285075541120023
step: 70, loss: 0.0010429604444652796
step: 80, loss: 0.0002667176886461675
step: 90, loss: 0.00014298531459644437
step: 100, loss: 0.00025088482652790844
step: 110, loss: 0.0032103420235216618
step: 120, loss: 0.026211176067590714
step: 130, loss: 0.0004580172826536
step: 140, loss: 0.0004333442193455994
step: 150, loss: 0.04638141766190529
step: 160, loss: 0.0017169073689728975
step: 170, loss: 0.0005377465859055519
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7677725118483412, f1=0.7972665148063781,
best_f1=0.7963386727688786
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```



```

packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
1916it [00:05, 321.73it/s]
load_f1 = 0.7769784172661871
real_f1 = 0.7868852459016392
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:18, 238.67it/s]

```

## 2.3 BASELINE DIRTY

### 2.3.1 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale              : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale              : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8195441365242004
step: 10, loss: 0.47407546639442444
step: 20, loss: 0.6177812218666077
step: 30, loss: 0.4729304015636444
step: 40, loss: 0.3245154023170471
step: 50, loss: 0.11152412742376328
step: 60, loss: 0.09638889133930206
step: 70, loss: 0.13652096688747406
step: 80, loss: 0.2841678559780121
step: 90, loss: 0.03560599684715271
step: 100, loss: 0.07416320592164993
step: 110, loss: 0.09793201088905334
step: 120, loss: 0.018986467272043228
step: 130, loss: 0.016724351793527603
step: 140, loss: 0.17376108467578888
step: 150, loss: 0.1278715431690216

```

```

step: 160, loss: 0.1802474558353424
step: 170, loss: 0.01602722518146038
step: 180, loss: 0.007160477805882692
step: 190, loss: 0.015178744681179523
step: 200, loss: 0.004203910939395428
step: 210, loss: 0.008384243585169315
step: 220, loss: 0.008161704987287521
step: 230, loss: 0.034652676433324814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9624724061810155, f1=0.9613259668508287,
best_f1=0.9613259668508287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03303317725658417
step: 10, loss: 0.0014675173442810774
step: 20, loss: 0.004178247414529324
step: 30, loss: 0.002264464972540736
step: 40, loss: 0.003322998061776161
step: 50, loss: 0.03974210470914841
step: 60, loss: 0.022072752937674522
step: 70, loss: 0.011695241555571556
step: 80, loss: 0.004441518802195787
step: 90, loss: 0.011772140860557556
step: 100, loss: 0.061304207891225815
step: 110, loss: 0.1513402760028839
step: 120, loss: 0.05238284170627594
step: 130, loss: 0.009110857732594013
step: 140, loss: 0.03420209512114525
step: 150, loss: 0.007590269669890404
step: 160, loss: 0.0043428996577858925
step: 170, loss: 0.01802080310881138
step: 180, loss: 0.007230069022625685
step: 190, loss: 0.17471514642238617
step: 200, loss: 0.0076883393339812756
step: 210, loss: 0.11048536002635956
step: 220, loss: 0.00379832717590034
step: 230, loss: 0.049254752695560455
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 2: dev\_f1=0.9652855543113102, f1=0.9720044792833147,

best\_f1=0.9720044792833147

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.050123341381549835

step: 10, loss: 0.013199817389249802

step: 20, loss: 0.006340948399156332

step: 30, loss: 0.025799265131354332

step: 40, loss: 0.020191779360175133

step: 50, loss: 0.00452136667445302

step: 60, loss: 0.015602873638272285

step: 70, loss: 0.00246604741550982

step: 80, loss: 0.025322619825601578

step: 90, loss: 0.03056531399488449

step: 100, loss: 0.0017380242934450507

step: 110, loss: 0.007919824682176113

step: 120, loss: 0.005076356697827578

step: 130, loss: 0.0027310866862535477

step: 140, loss: 0.0012104605557397008

step: 150, loss: 0.0056067053228616714

step: 160, loss: 0.004743615631014109

step: 170, loss: 0.00976614560931921

step: 180, loss: 0.005065863952040672

step: 190, loss: 0.0025309554766863585

step: 200, loss: 0.005548481363803148

step: 210, loss: 0.006326686590909958

step: 220, loss: 0.009125233627855778

step: 230, loss: 0.03122362680733204

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 3: dev\_f1=0.9697648376259798, f1=0.967452300785634,

best\_f1=0.967452300785634

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.023392872884869576

step: 10, loss: 0.0020451005548238754

step: 20, loss: 0.0017084423452615738

step: 30, loss: 0.0006496899295598269

```

step: 40, loss: 0.010971179232001305
step: 50, loss: 0.002349646994844079
step: 60, loss: 0.0026380938943475485
step: 70, loss: 0.008618253283202648
step: 80, loss: 0.010603527538478374
step: 90, loss: 0.006234285421669483
step: 100, loss: 0.008531213738024235
step: 110, loss: 0.02284897305071354
step: 120, loss: 0.004158053081482649
step: 130, loss: 0.0019377995049580932
step: 140, loss: 0.0004179902607575059
step: 150, loss: 0.0010347666684538126
step: 160, loss: 0.0015095019480213523
step: 170, loss: 0.0013549529248848557
step: 180, loss: 0.09380220621824265
step: 190, loss: 0.010667501017451286
step: 200, loss: 0.0041187116876244545
step: 210, loss: 0.0007740950095467269
step: 220, loss: 0.0025533498264849186
step: 230, loss: 0.001098055043257773
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9709821428571428, f1=0.9708520179372198,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011377283371984959
step: 10, loss: 0.001996650593355298
step: 20, loss: 0.0013101372169330716
step: 30, loss: 0.000385490246117115
step: 40, loss: 0.0007943066884763539
step: 50, loss: 0.0014428584836423397
step: 60, loss: 0.0023683574981987476
step: 70, loss: 0.00032996677327901125
step: 80, loss: 0.0015672557055950165
step: 90, loss: 0.00961505901068449
step: 100, loss: 0.00871372316032648
step: 110, loss: 0.0014833511086180806
step: 120, loss: 0.0408850759267807
step: 130, loss: 0.0176126416772604
step: 140, loss: 0.0014956833329051733
step: 150, loss: 0.0001954599138116464

```

```

step: 160, loss: 0.020008835941553116
step: 170, loss: 0.007948321290314198
step: 180, loss: 0.0011594267562031746
step: 190, loss: 0.00037172934389673173
step: 200, loss: 0.006311663892120123
step: 210, loss: 0.007685631513595581
step: 220, loss: 0.002565454225987196
step: 230, loss: 0.0017287801019847393
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9743589743589743, f1=0.9698324022346367,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032401077914983034
step: 10, loss: 0.0006363061256706715
step: 20, loss: 0.0008597509004175663
step: 30, loss: 0.005128677003085613
step: 40, loss: 0.0005562198348343372
step: 50, loss: 0.006074873730540276
step: 60, loss: 0.0008218616130761802
step: 70, loss: 0.03658001869916916
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.002117460360750556
step: 90, loss: 0.0002937423123512417
step: 100, loss: 0.0007012897403910756
step: 110, loss: 0.045261185616254807
step: 120, loss: 0.00026442843955010176
step: 130, loss: 0.00018926015764009207
step: 140, loss: 0.0012478586286306381
step: 150, loss: 0.0003612050204537809
step: 160, loss: 0.000818470143713057
step: 170, loss: 0.0044377585873007774
step: 180, loss: 0.0012723825639113784
step: 190, loss: 0.0010915766470134258
step: 200, loss: 0.03433637693524361
step: 210, loss: 0.0012807163875550032
step: 220, loss: 0.00040710510802455246
step: 230, loss: 0.0008303755894303322
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 6: dev_f1=0.970917225950783, f1=0.9730337078651685,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015597093850374222
step: 10, loss: 0.0025524436496198177
step: 20, loss: 0.0006967289955355227
step: 30, loss: 0.004243372939527035
step: 40, loss: 0.000828475400339812
step: 50, loss: 0.0006963500054553151
step: 60, loss: 0.12144123017787933
step: 70, loss: 0.0002705845981836319
step: 80, loss: 0.00017342530190944672
step: 90, loss: 0.00023029433214105666
step: 100, loss: 0.0006641870131716132
step: 110, loss: 0.015009772963821888
step: 120, loss: 0.0008834770997054875
step: 130, loss: 0.0002805653493851423
step: 140, loss: 0.00016628330922685564
step: 150, loss: 0.0001829165412345901
step: 160, loss: 0.00031036583823151886
step: 170, loss: 0.00026215208345092833
step: 180, loss: 0.014398383907973766
step: 190, loss: 0.00012039014109177515
step: 200, loss: 0.0007931966101750731
step: 210, loss: 0.0003010607906617224
step: 220, loss: 0.00727933319285512
step: 230, loss: 0.014910245314240456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9659090909090909, f1=0.9636363636363636,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006794249638915062
step: 10, loss: 0.0007258045370690525
step: 20, loss: 0.0014494959032163024
step: 30, loss: 0.0004952256567776203
step: 40, loss: 0.0003360325936228037
step: 50, loss: 0.00034033923293463886
step: 60, loss: 0.00021590841060969979
step: 70, loss: 0.002660430269315839
step: 80, loss: 0.0007665416342206299
step: 90, loss: 0.0002791301521938294
step: 100, loss: 0.00018417148385196924

```

```

step: 110, loss: 0.00023326848167926073
step: 120, loss: 0.000372917769709602
step: 130, loss: 0.008905614726245403
step: 140, loss: 0.00012550388055387884
step: 150, loss: 0.000894546159543097
step: 160, loss: 0.0030933290254324675
step: 170, loss: 0.001800323254428804
step: 180, loss: 0.002155660418793559
step: 190, loss: 0.00012336716463323683
step: 200, loss: 0.013458115980029106
step: 210, loss: 9.956646681530401e-05
step: 220, loss: 0.00014378075138665736
step: 230, loss: 0.0005344516830518842
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.972129319955407, f1=0.9741863075196409,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001352046529063955
step: 10, loss: 0.00019504540250636637
step: 20, loss: 0.0012169064721092582
step: 30, loss: 0.0004630167386494577
step: 40, loss: 0.00012083523324690759
step: 50, loss: 0.0001345098135061562
step: 60, loss: 0.0006599624175578356
step: 70, loss: 0.0004934785538353026
step: 80, loss: 0.03331120312213898
step: 90, loss: 0.012503380887210369
step: 100, loss: 0.00033715926110744476
step: 110, loss: 0.00011904593702638522
step: 120, loss: 0.014725885353982449
step: 130, loss: 0.00011797384649980813
step: 140, loss: 0.0003663214738480747
step: 150, loss: 7.195158104877919e-05
step: 160, loss: 0.0016445894725620747
step: 170, loss: 0.00013590382877737284
step: 180, loss: 0.0009998311288654804
step: 190, loss: 0.00012274057371541858
step: 200, loss: 0.0002584758913144469
step: 210, loss: 0.0003099343157373369
step: 220, loss: 0.0071748774498701096
step: 230, loss: 0.00019842464826069772
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```



```

epoch 9: dev_f1=0.9722530521642618, f1=0.9732739420935412,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018338611698709428
step: 10, loss: 9.852766379481182e-05
step: 20, loss: 0.0003381922433618456
step: 30, loss: 0.00012266664998605847
step: 40, loss: 4.290114520699717e-05
step: 50, loss: 0.00037845392944291234
step: 60, loss: 0.06804771721363068
step: 70, loss: 0.000220855901716277
step: 80, loss: 0.00014267858932726085
step: 90, loss: 9.20703896554187e-05
step: 100, loss: 0.0024558936711400747
step: 110, loss: 0.00015298354264814407
step: 120, loss: 0.041084736585617065
step: 130, loss: 9.535066783428192e-05
step: 140, loss: 0.0027999994345009327
step: 150, loss: 0.0002011773904087022
step: 160, loss: 0.0001021583957481198
step: 170, loss: 0.0002835017803590745
step: 180, loss: 0.00033728708513081074
step: 190, loss: 9.008110646391287e-05
step: 200, loss: 7.037644536467269e-05
step: 210, loss: 9.376084926771e-05
step: 220, loss: 0.0001296273694606498
step: 230, loss: 0.0007802998297847807
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9708520179372198, f1=0.9718785151856018,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.641484262421727e-05
step: 10, loss: 8.236946450779215e-05
step: 20, loss: 5.2134815632598475e-05
step: 30, loss: 0.00011866707063745707
step: 40, loss: 0.0005106423050165176
step: 50, loss: 0.00034301739651709795
step: 60, loss: 7.221222040243447e-05
step: 70, loss: 0.0001557640207465738
step: 80, loss: 0.000191106169950217
step: 90, loss: 8.575761603424326e-05
step: 100, loss: 8.553605584893376e-05

```

```

step: 110, loss: 0.002955008065328002
step: 120, loss: 0.00044912536395713687
step: 130, loss: 9.279629011871293e-05
step: 140, loss: 7.84333678893745e-05
step: 150, loss: 5.198598591960035e-05
step: 160, loss: 5.021307879360393e-05
step: 170, loss: 0.000669079483486712
step: 180, loss: 9.052402310771868e-05
step: 190, loss: 9.70638866419904e-05
step: 200, loss: 8.519194670952857e-05
step: 210, loss: 5.910828986088745e-05
step: 220, loss: 8.200043521355838e-05
step: 230, loss: 0.00020602990116458386
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.972129319955407, f1=0.9730337078651685,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014588600606657565
step: 10, loss: 0.0001422012283001095
step: 20, loss: 4.542344686342403e-05
step: 30, loss: 9.846915781963617e-05
step: 40, loss: 5.315943781170063e-05
step: 50, loss: 5.294725997373462e-05
step: 60, loss: 0.009845009073615074
step: 70, loss: 6.498287984868512e-05
step: 80, loss: 3.9630423998460174e-05
step: 90, loss: 5.776122270617634e-05
step: 100, loss: 4.2416941141709685e-05
step: 110, loss: 7.154548802645877e-05
step: 120, loss: 9.112884436035529e-05
step: 130, loss: 8.789795538177714e-05
step: 140, loss: 4.63100805063732e-05
step: 150, loss: 4.6803299483144656e-05
step: 160, loss: 0.0002500223636161536
step: 170, loss: 6.963771011214703e-05
step: 180, loss: 4.7669447667431086e-05
step: 190, loss: 6.116086296970025e-05
step: 200, loss: 0.01063668355345726
step: 210, loss: 5.395720654632896e-05
step: 220, loss: 0.0004098965146113187
step: 230, loss: 0.00021748902508988976
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 12: dev_f1=0.9698324022346367, f1=0.971815107102593,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.079319089418277e-05
step: 10, loss: 4.916265243082307e-05
step: 20, loss: 7.526845001848415e-05
step: 30, loss: 0.008261105976998806
step: 40, loss: 0.022884061560034752
step: 50, loss: 6.032712190062739e-05
step: 60, loss: 3.6979276046622545e-05
step: 70, loss: 6.633603334194049e-05
step: 80, loss: 0.030366986989974976
step: 90, loss: 3.3522635931149125e-05
step: 100, loss: 0.00013982653035782278
step: 110, loss: 4.896277096122503e-05
step: 120, loss: 4.517805791692808e-05
step: 130, loss: 0.00011174703104188666
step: 140, loss: 0.0010811852989718318
step: 150, loss: 0.0004408960521686822
step: 160, loss: 0.0015683395322412252
step: 170, loss: 5.6140099331969395e-05
step: 180, loss: 6.091199247748591e-05
step: 190, loss: 3.0092029192019254e-05
step: 200, loss: 4.474941670196131e-05
step: 210, loss: 5.836726995767094e-05
step: 220, loss: 5.825240077683702e-05
step: 230, loss: 3.9761602238286287e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9733924611973392, f1=0.9743016759776536,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.8867299483390525e-05
step: 10, loss: 2.5744850063347258e-05
step: 20, loss: 0.0002033481578109786
step: 30, loss: 0.00010818502050824463
step: 40, loss: 3.0311526643345132e-05
step: 50, loss: 9.320558456238359e-05
step: 60, loss: 5.3143277909839526e-05
step: 70, loss: 4.1776878788368776e-05
step: 80, loss: 7.060572534101084e-05
step: 90, loss: 0.00012520306336227804
step: 100, loss: 0.17306974530220032

```

```

step: 110, loss: 7.19861636753194e-05
step: 120, loss: 0.0019128115382045507
step: 130, loss: 5.957532994216308e-05
step: 140, loss: 4.259961860952899e-05
step: 150, loss: 6.366049638018012e-05
step: 160, loss: 3.112335252808407e-05
step: 170, loss: 0.00016081788635347039
step: 180, loss: 5.375983892008662e-05
step: 190, loss: 7.657564128749073e-05
step: 200, loss: 4.205291043035686e-05
step: 210, loss: 8.636930579086766e-05
step: 220, loss: 5.9084257372887805e-05
step: 230, loss: 2.5409588488400914e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9723145071982282, f1=0.9731543624161074,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.323294706409797e-05
step: 10, loss: 6.535763532156125e-05
step: 20, loss: 4.804586205864325e-05
step: 30, loss: 8.234873530454934e-05
step: 40, loss: 6.208041304489598e-05
step: 50, loss: 6.324637070065364e-05
step: 60, loss: 0.0003820264246314764
step: 70, loss: 8.369533315999433e-05
step: 80, loss: 0.0001083058668882586
step: 90, loss: 4.310150325181894e-05
step: 100, loss: 4.9323902203468606e-05
step: 110, loss: 5.681303809978999e-05
step: 120, loss: 6.02578729740344e-05
step: 130, loss: 0.00014233877300284803
step: 140, loss: 0.003722880966961384
step: 150, loss: 0.00014673714758828282
step: 160, loss: 0.0007518968195654452
step: 170, loss: 3.8744834455428645e-05
step: 180, loss: 4.8898458771873266e-05
step: 190, loss: 7.403765630442649e-05
step: 200, loss: 4.338949656812474e-05
step: 210, loss: 0.01855967752635479
step: 220, loss: 8.270599937532097e-05
step: 230, loss: 5.92210708418861e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```
epoch 15: dev_f1=0.971815107102593, f1=0.9728506787330317,
best_f1=0.9698324022346367
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.
```

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
2473it [00:11, 223.64it/s]
load_f1 = 0.9754464285714286
real_f1 = 0.9732739420935412
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
4398it [00:18, 243.22it/s]
```

### 2.3.2 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
```

```

!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Dirty/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Dirty/DBLP-GoogleScholar \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8057878017425537
step: 10, loss: 0.404278039932251
step: 20, loss: 0.4881853461265564
step: 30, loss: 0.4338069260120392
step: 40, loss: 0.2872839868068695
step: 50, loss: 0.1384032964706421
step: 60, loss: 0.12388639897108078
step: 70, loss: 0.11177825182676315
step: 80, loss: 0.33856481313705444
step: 90, loss: 0.13579367101192474
step: 100, loss: 0.3757425546646118
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 110, loss: 0.026056647300720215
step: 120, loss: 0.054528314620256424
step: 130, loss: 0.039365630596876144
step: 140, loss: 0.06931811571121216
step: 150, loss: 0.10455036908388138
step: 160, loss: 0.14284799993038177
step: 170, loss: 0.13195350766181946
step: 180, loss: 0.10765492916107178
step: 190, loss: 0.03010805882513523
step: 200, loss: 0.11617182195186615
step: 210, loss: 0.09328713268041611
step: 220, loss: 0.07379496097564697
step: 230, loss: 0.06078560650348663
step: 240, loss: 0.06755766272544861
step: 250, loss: 0.03453412652015686
step: 260, loss: 0.0319848395884037
step: 270, loss: 0.01708262413740158
step: 280, loss: 0.10481993108987808
step: 290, loss: 0.055579498410224915
step: 300, loss: 0.1918923556804657
step: 310, loss: 0.060280147939920425
step: 320, loss: 0.027984747663140297
step: 330, loss: 0.1355995535850525

```

```

step: 340, loss: 0.14845536649227142
step: 350, loss: 0.10590236634016037
step: 360, loss: 0.04436859115958214
step: 370, loss: 0.06812230497598648
step: 380, loss: 0.12826108932495117
step: 390, loss: 0.011358575895428658
step: 400, loss: 0.011700578033924103
step: 410, loss: 0.041735414415597916
step: 420, loss: 0.023932568728923798
step: 430, loss: 0.07277039438486099
step: 440, loss: 0.11193355917930603
step: 450, loss: 0.0715416669845581
step: 460, loss: 0.17188507318496704
step: 470, loss: 0.2380518615245819
step: 480, loss: 0.24508944153785706
step: 490, loss: 0.03245210647583008
step: 500, loss: 0.04597184434533119
step: 510, loss: 0.08414476364850998
step: 520, loss: 0.07926616817712784
step: 530, loss: 0.06784297525882721
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9264029438822446, f1=0.9318701417466849,
best_f1=0.9318701417466849
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14915576577186584
step: 10, loss: 0.10307631641626358
step: 20, loss: 0.06619442254304886
step: 30, loss: 0.0845242440700531
step: 40, loss: 0.009772928431630135
step: 50, loss: 0.03656359389424324
step: 60, loss: 0.11815040558576584
step: 70, loss: 0.2085760235786438
step: 80, loss: 0.06605149060487747
step: 90, loss: 0.01671655662357807
step: 100, loss: 0.3259745240211487
step: 110, loss: 0.024450521916151047
step: 120, loss: 0.06495653837919235
step: 130, loss: 0.009507761336863041
step: 140, loss: 0.03464021906256676
step: 150, loss: 0.018982453271746635

```



```

step: 160, loss: 0.013304634019732475
step: 170, loss: 0.06132826209068298
step: 180, loss: 0.005711632315069437
step: 190, loss: 0.022775691002607346
step: 200, loss: 0.015033924952149391
step: 210, loss: 0.007794907782226801
step: 220, loss: 0.10622857511043549
step: 230, loss: 0.01819000393152237
step: 240, loss: 0.09913261979818344
step: 250, loss: 0.07442431896924973
step: 260, loss: 0.07486361265182495
step: 270, loss: 0.18348243832588196
step: 280, loss: 0.19975817203521729
step: 290, loss: 0.0827171579003334
step: 300, loss: 0.0416659377515316
step: 310, loss: 0.061631690710783005
step: 320, loss: 0.2521200478076935
step: 330, loss: 0.043052565306425095
step: 340, loss: 0.0028078295290470123
step: 350, loss: 0.03038780577480793
step: 360, loss: 0.031123211607336998
step: 370, loss: 0.004098028410226107
step: 380, loss: 0.11232870817184448
step: 390, loss: 0.04008350148797035
step: 400, loss: 0.024025702849030495
step: 410, loss: 0.0005193281685933471
step: 420, loss: 0.05709375813603401
step: 430, loss: 0.024066418409347534
step: 440, loss: 0.008039616979658604
step: 450, loss: 0.05441313982009888
step: 460, loss: 0.24698856472969055
step: 470, loss: 0.06146882474422455
step: 480, loss: 0.17292508482933044
step: 490, loss: 0.032510578632354736
step: 500, loss: 0.007867427542805672
step: 510, loss: 0.03890003263950348
step: 520, loss: 0.10711177438497543
step: 530, loss: 0.048895224928855896
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.935483870967742, f1=0.933774834437086,
best_f1=0.933774834437086
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.013273433782160282
step: 10, loss: 0.04304787516593933
step: 20, loss: 0.2227240949869156
step: 30, loss: 0.13191962242126465
step: 40, loss: 0.008550678379833698
step: 50, loss: 0.009736898355185986
step: 60, loss: 0.00210277340374887
step: 70, loss: 0.04118962585926056
step: 80, loss: 0.0020596208050847054
step: 90, loss: 0.09268073737621307
step: 100, loss: 0.153419628739357
step: 110, loss: 0.03046480379998684
step: 120, loss: 0.010649123229086399
step: 130, loss: 0.013987586833536625
step: 140, loss: 0.00973983108997345
step: 150, loss: 0.008337512612342834
step: 160, loss: 0.0017758923349902034
step: 170, loss: 0.009610000066459179
step: 180, loss: 0.003227761248126626
step: 190, loss: 0.024137595668435097
step: 200, loss: 0.033095989376306534
step: 210, loss: 0.06430456787347794
step: 220, loss: 0.004668320529162884
step: 230, loss: 0.049712348729372025
step: 240, loss: 0.012587714940309525
step: 250, loss: 0.009338521398603916
step: 260, loss: 0.0056874570436775684
step: 270, loss: 0.005499666091054678
step: 280, loss: 0.012102898210287094
step: 290, loss: 0.08376724272966385
step: 300, loss: 0.015052738599479198
step: 310, loss: 0.20098081231117249
step: 320, loss: 0.16503717005252838
step: 330, loss: 0.007473729085177183
step: 340, loss: 0.003833093447610736
step: 350, loss: 0.025910893455147743
step: 360, loss: 0.017047859728336334
step: 370, loss: 0.003361013950780034
step: 380, loss: 0.07835496962070465
step: 390, loss: 0.007413353770971298
step: 400, loss: 0.009597284719347954
step: 410, loss: 0.013063809834420681
step: 420, loss: 0.017766976729035378
step: 430, loss: 0.070820152759552
step: 440, loss: 0.14634010195732117
step: 450, loss: 0.14394699037075043
```

```

step: 460, loss: 0.029161740094423294
step: 470, loss: 0.004214240238070488
step: 480, loss: 0.001729035982862115
step: 490, loss: 0.025001663714647293
step: 500, loss: 0.15126988291740417
step: 510, loss: 0.0027342387475073338
step: 520, loss: 0.004030552227050066
step: 530, loss: 0.1001313254237175
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9368954398894519, f1=0.9381918819188192,
best_f1=0.9381918819188192
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0030738450586795807
step: 10, loss: 0.006365463603287935
step: 20, loss: 0.019351646304130554
step: 30, loss: 0.0008460519602522254
step: 40, loss: 0.0013390115927904844
step: 50, loss: 0.006811054889112711
step: 60, loss: 0.00028689089231193066
step: 70, loss: 0.0014090919867157936
step: 80, loss: 0.030802445486187935
step: 90, loss: 0.023393627256155014
step: 100, loss: 0.001900530536659062
step: 110, loss: 0.0004282141162548214
step: 120, loss: 0.021058961749076843
step: 130, loss: 0.000975720351561904
step: 140, loss: 0.010763084515929222
step: 150, loss: 0.012259507551789284
step: 160, loss: 0.001544415601529181
step: 170, loss: 0.005694353021681309
step: 180, loss: 0.09243201464414597
step: 190, loss: 0.04527529701590538
step: 200, loss: 0.0057679410092532635
step: 210, loss: 0.10351160168647766
step: 220, loss: 0.00682542659342289
step: 230, loss: 0.21339938044548035
step: 240, loss: 0.004585194867104292
step: 250, loss: 0.12110678106546402
step: 260, loss: 0.17873461544513702
step: 270, loss: 0.02915334701538086

```

```

step: 280, loss: 0.004644281230866909
step: 290, loss: 0.16839568316936493
step: 300, loss: 0.0022532015573233366
step: 310, loss: 0.004227194003760815
step: 320, loss: 0.027871932834386826
step: 330, loss: 0.04853206127882004
step: 340, loss: 0.029049277305603027
step: 350, loss: 0.006153805647045374
step: 360, loss: 0.013115433976054192
step: 370, loss: 0.006787713151425123
step: 380, loss: 0.0013041761703789234
step: 390, loss: 0.14029407501220703
step: 400, loss: 0.011870818212628365
step: 410, loss: 0.021359674632549286
step: 420, loss: 0.037370361387729645
step: 430, loss: 0.009581523947417736
step: 440, loss: 0.026516135782003403
step: 450, loss: 0.005571072455495596
step: 460, loss: 0.001688013318926096
step: 470, loss: 0.0036977699492126703
step: 480, loss: 0.001023069373331964
step: 490, loss: 0.03291589021682739
step: 500, loss: 0.007474142592400312
step: 510, loss: 0.00685807503759861
step: 520, loss: 0.06806375086307526
step: 530, loss: 0.00350599130615592
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9366494603472549, f1=0.9270346117867166,
best_f1=0.9381918819188192
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019702566787600517
step: 10, loss: 0.02864585816860199
step: 20, loss: 0.0008151044021360576
step: 30, loss: 0.004755102097988129
step: 40, loss: 0.04773765802383423
step: 50, loss: 0.01106580812484026
step: 60, loss: 0.002079347148537636
step: 70, loss: 0.0024313400499522686
step: 80, loss: 0.0011099543189629912
step: 90, loss: 0.007027193438261747
step: 100, loss: 0.0013460888294503093
step: 110, loss: 0.0009196166065521538
step: 120, loss: 0.003812713548541069
step: 130, loss: 0.015642091631889343

```

```

step: 140, loss: 0.0402960442006588
step: 150, loss: 0.01013888604938984
step: 160, loss: 0.18704698979854584
step: 170, loss: 0.009016241878271103
step: 180, loss: 0.0008837728528305888
step: 190, loss: 0.0004706111503764987
step: 200, loss: 0.0006822245777584612
step: 210, loss: 0.0012374374782666564
step: 220, loss: 0.0005158102139830589
step: 230, loss: 0.0010566309792920947
step: 240, loss: 0.014016413129866123
step: 250, loss: 0.026363415643572807
step: 260, loss: 0.0013369375374168158
step: 270, loss: 0.0068876962177455425
step: 280, loss: 0.06672799587249756
step: 290, loss: 0.0012696725316345692
step: 300, loss: 0.025096306577324867
step: 310, loss: 0.03690765053033829
step: 320, loss: 0.00713652279227972
step: 330, loss: 0.0005447315634228289
step: 340, loss: 0.002042618114501238
step: 350, loss: 0.04613924399018288
step: 360, loss: 0.0006929748924449086
step: 370, loss: 0.006958496756851673
step: 380, loss: 0.11524651199579239
step: 390, loss: 0.002037331461906433
step: 400, loss: 0.015053866431117058
step: 410, loss: 0.0006310395547188818
step: 420, loss: 0.010900603607296944
step: 430, loss: 0.00040471157990396023
step: 440, loss: 0.0038352699484676123
step: 450, loss: 0.0005093790241517127
step: 460, loss: 0.0005826037959195673
step: 470, loss: 0.001851142500527203
step: 480, loss: 0.0010181303368881345
step: 490, loss: 0.0013458210742101073
step: 500, loss: 0.001411996316164732
step: 510, loss: 0.005566912703216076
step: 520, loss: 0.00027617637533694506
step: 530, loss: 0.015349420718848705
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9392111368909513, f1=0.9308584686774942,

```

```
best_f1=0.9308584686774942
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006428857333958149
step: 10, loss: 0.009248154237866402
step: 20, loss: 0.002594602759927511
step: 30, loss: 0.004773769993335009
step: 40, loss: 0.0007929482380859554
step: 50, loss: 0.0025908625684678555
step: 60, loss: 0.0013669101754203439
step: 70, loss: 0.003017353592440486
step: 80, loss: 0.0012396727688610554
step: 90, loss: 0.002683187136426568
step: 100, loss: 0.15946656465530396
step: 110, loss: 0.02481784299015999
step: 120, loss: 0.0004562684625852853
step: 130, loss: 0.00032824103254824877
step: 140, loss: 0.0021221316419541836
step: 150, loss: 0.008035572245717049
step: 160, loss: 0.00015809810429345816
step: 170, loss: 0.0002965217863675207
step: 180, loss: 0.0006980579346418381
step: 190, loss: 0.004392832983285189
step: 200, loss: 0.0001261386350961402
step: 210, loss: 0.00016478188626933843
step: 220, loss: 0.003379503730684519
step: 230, loss: 0.0001528807042632252
step: 240, loss: 0.00023403890372719616
step: 250, loss: 0.00012384958972688764
step: 260, loss: 0.00021719337382819504
step: 270, loss: 0.0001334100088570267
step: 280, loss: 0.002884226618334651
step: 290, loss: 0.002199731534346938
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 300, loss: 0.0007805594359524548
step: 310, loss: 0.0040116021409630775
step: 320, loss: 0.036405887454748154
step: 330, loss: 0.004101363476365805
step: 340, loss: 0.0031188209541141987
step: 350, loss: 0.026428351178765297
step: 360, loss: 0.0011620186269283295
step: 370, loss: 0.006302646826952696
step: 380, loss: 0.003601477947086096
step: 390, loss: 0.04377108812332153
step: 400, loss: 0.0007538556237705052
step: 410, loss: 0.0007392630213871598
step: 420, loss: 0.0007138191722333431
```

```

step: 430, loss: 0.0012709207367151976
step: 440, loss: 0.0013768490171059966
step: 450, loss: 0.0009448014316149056
step: 460, loss: 0.03174778074026108
step: 470, loss: 0.008777277544140816
step: 480, loss: 0.019727172330021858
step: 490, loss: 0.005836464464664459
step: 500, loss: 0.004472211468964815
step: 510, loss: 0.00026242094463668764
step: 520, loss: 0.0006859356653876603
step: 530, loss: 0.007540483959019184
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9324515824279641, f1=0.9195294117647058,
best_f1=0.9308584686774942
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08772934973239899
step: 10, loss: 0.0005963091389276087
step: 20, loss: 0.013966280966997147
step: 30, loss: 0.0010940806241706014
step: 40, loss: 0.000212117112823762
step: 50, loss: 0.013188117183744907
step: 60, loss: 0.031899452209472656
step: 70, loss: 0.013207484036684036
step: 80, loss: 0.007375212386250496
step: 90, loss: 0.00020254771516192704
step: 100, loss: 0.002887090900912881
step: 110, loss: 0.007534723728895187
step: 120, loss: 0.00021239754278212786
step: 130, loss: 8.11016361694783e-05
step: 140, loss: 0.0006474986439570785
step: 150, loss: 7.09478190401569e-05
step: 160, loss: 0.0008794025634415448
step: 170, loss: 0.0018486297922208905
step: 180, loss: 8.997073746286333e-05
step: 190, loss: 0.0017999822739511728
step: 200, loss: 0.00018946484487969428
step: 210, loss: 0.005524679087102413
step: 220, loss: 0.00011730643745977432
step: 230, loss: 0.0016670533223077655
step: 240, loss: 0.004963277839124203
step: 250, loss: 0.00021788626327179372
step: 260, loss: 0.05496610328555107
step: 270, loss: 4.131526293349452e-05
step: 280, loss: 0.06888218224048615

```

```

step: 290, loss: 0.0013112393207848072
step: 300, loss: 0.0008026604191400111
step: 310, loss: 0.0008433415787294507
step: 320, loss: 0.02395595982670784
step: 330, loss: 0.00026300185709260404
step: 340, loss: 0.0017987923929467797
step: 350, loss: 0.003342230571433902
step: 360, loss: 0.0035879435017704964
step: 370, loss: 0.007804930210113525
step: 380, loss: 0.0019470949191600084
step: 390, loss: 0.0005225017666816711
step: 400, loss: 7.907988037914038e-05
step: 410, loss: 0.00042066595051437616
step: 420, loss: 0.00010106246918439865
step: 430, loss: 9.29079542402178e-05
step: 440, loss: 7.20144307706505e-05
step: 450, loss: 0.00028617962379939854
step: 460, loss: 0.0031496225856244564
step: 470, loss: 0.0008527713362127542
step: 480, loss: 0.002968801185488701
step: 490, loss: 0.0006904836045578122
step: 500, loss: 0.00019378226716071367
step: 510, loss: 0.0009547746158204973
step: 520, loss: 0.00012916287232656032
step: 530, loss: 0.00015491846716031432
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9403606102635228, f1=0.9365446966188051,
best_f1=0.9365446966188051
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000633028510492295
step: 10, loss: 0.0007674224907532334
step: 20, loss: 0.031447988003492355
step: 30, loss: 0.00014292003470472991
step: 40, loss: 0.00012101404718123376
step: 50, loss: 0.00012187330867163837
step: 60, loss: 0.0004458562470972538
step: 70, loss: 0.00020236163982190192
step: 80, loss: 0.00010619031672831625
step: 90, loss: 6.79295408190228e-05
step: 100, loss: 0.0007421920308843255

```



```

step: 110, loss: 0.003154180943965912
step: 120, loss: 0.015216155909001827
step: 130, loss: 0.0025388705544173717
step: 140, loss: 5.15041065227706e-05
step: 150, loss: 0.0036522946320474148
step: 160, loss: 0.0002655287680681795
step: 170, loss: 0.00017058140656445175
step: 180, loss: 0.0022747903130948544
step: 190, loss: 0.012502527795732021
step: 200, loss: 0.0002650651440490037
step: 210, loss: 0.00016670030890963972
step: 220, loss: 0.0007968789432197809
step: 230, loss: 0.0005705623188987374
step: 240, loss: 0.00020622271404135972
step: 250, loss: 0.012615809217095375
step: 260, loss: 0.0044689299538731575
step: 270, loss: 4.4906733819516376e-05
step: 280, loss: 0.0182805098593235
step: 290, loss: 0.011602871119976044
step: 300, loss: 6.797110836487263e-05
step: 310, loss: 0.005254289600998163
step: 320, loss: 0.00014802312944084406
step: 330, loss: 0.011916938237845898
step: 340, loss: 0.0010947496630251408
step: 350, loss: 0.001352313905954361
step: 360, loss: 0.0001953202299773693
step: 370, loss: 0.0004191052867099643
step: 380, loss: 0.009954044595360756
step: 390, loss: 0.00024201307678595185
step: 400, loss: 0.0028459331952035427
step: 410, loss: 7.749510405119509e-05
step: 420, loss: 6.529329402837902e-05
step: 430, loss: 0.0022141234949231148
step: 440, loss: 0.0003705744748003781
step: 450, loss: 0.001107282587327063
step: 460, loss: 0.00037880599847994745
step: 470, loss: 0.00024741023662500083
step: 480, loss: 0.0003872651723213494
step: 490, loss: 5.376643093768507e-05
step: 500, loss: 9.681105439085513e-05
step: 510, loss: 4.7363864723592997e-05
step: 520, loss: 0.00011893034388776869
step: 530, loss: 0.0001973543403437361
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or

```

```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.9418874941887495, f1=0.9346826126954921,
best_f1=0.9346826126954921
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037970070843584836
step: 10, loss: 7.230293704196811e-05
step: 20, loss: 5.6919325288617983e-05
step: 30, loss: 0.00013953661255072802
step: 40, loss: 0.00022976365289650857
step: 50, loss: 0.00016468479589093477
step: 60, loss: 0.0002107276231981814
step: 70, loss: 0.11149980872869492
step: 80, loss: 0.00032105858554132283
step: 90, loss: 0.0033429318573325872
step: 100, loss: 0.00031528339604847133
step: 110, loss: 5.5457509006373584e-05
step: 120, loss: 5.592342859017663e-05
step: 130, loss: 4.8965193855110556e-05
step: 140, loss: 0.03942713886499405
step: 150, loss: 0.008077445439994335
step: 160, loss: 0.0009321096586063504
step: 170, loss: 9.498734289081767e-05
step: 180, loss: 5.5622778745600954e-05
step: 190, loss: 0.00011964188888669014
step: 200, loss: 0.0044811624102294445
step: 210, loss: 6.208132981555536e-05
step: 220, loss: 0.00019472691928967834
step: 230, loss: 4.140007513342425e-05
step: 240, loss: 0.00027898960979655385
step: 250, loss: 0.000304417364532128
step: 260, loss: 0.007773308083415031
step: 270, loss: 0.00016828189836815
step: 280, loss: 9.081915050046518e-05
step: 290, loss: 4.776194327860139e-05
step: 300, loss: 0.00029797464958392084
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 310, loss: 8.745321247261018e-05
step: 320, loss: 0.00021409244800452143
step: 330, loss: 0.00015071961388457566
step: 340, loss: 0.00017618807032704353
step: 350, loss: 4.004796574008651e-05
step: 360, loss: 0.050501950085163116
step: 370, loss: 0.00139134400524199
step: 380, loss: 3.208129055565223e-05
step: 390, loss: 3.654287138488144e-05

```

```

step: 400, loss: 0.0005982090369798243
step: 410, loss: 0.0015890925424173474
step: 420, loss: 0.00920687336474657
step: 430, loss: 0.0018293879693374038
step: 440, loss: 0.0003088241792283952
step: 450, loss: 0.005643579177558422
step: 460, loss: 0.0009734797640703619
step: 470, loss: 0.00037834892282262444
step: 480, loss: 0.002046093577519059
step: 490, loss: 0.005766879767179489
step: 500, loss: 0.0014022956602275372
step: 510, loss: 0.0013445541262626648
step: 520, loss: 0.0001929115824168548
step: 530, loss: 0.001294954214245081
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.9421028253821214, f1=0.9352319706017456,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.759185195434839e-05
step: 10, loss: 2.2463154891738668e-05
step: 20, loss: 8.254881686298177e-05
step: 30, loss: 7.070771243888885e-05
step: 40, loss: 0.00027543253963813186
step: 50, loss: 0.00034552474971860647
step: 60, loss: 0.00014113170618657023
step: 70, loss: 0.0017705165082588792
step: 80, loss: 0.0006065102643333375
step: 90, loss: 0.0017214362742379308
step: 100, loss: 6.955805292818695e-05
step: 110, loss: 0.0002456982037983835
step: 120, loss: 5.172595410840586e-05
step: 130, loss: 4.342581451055594e-05
step: 140, loss: 0.0002258900785818696
step: 150, loss: 0.0009804421570152044
step: 160, loss: 0.001970831537619233
step: 170, loss: 0.003820391371846199
step: 180, loss: 0.04313065484166145
step: 190, loss: 8.39844869915396e-05
step: 200, loss: 9.170564590021968e-05
step: 210, loss: 5.277944728732109e-05

```

```

step: 220, loss: 6.077981015550904e-05
step: 230, loss: 0.00019126562983728945
step: 240, loss: 0.00015035334217827767
step: 250, loss: 0.001170849776826799
step: 260, loss: 0.001547646475955844
step: 270, loss: 7.645356527063996e-05
step: 280, loss: 9.891360241454095e-05
step: 290, loss: 5.7298082538181916e-05
step: 300, loss: 9.581758786225691e-05
step: 310, loss: 6.108342495281249e-05
step: 320, loss: 7.71493578213267e-05
step: 330, loss: 0.004013366531580687
step: 340, loss: 3.697900683619082e-05
step: 350, loss: 0.0002109541092067957
step: 360, loss: 6.609028787352145e-05
step: 370, loss: 0.0007339751464314759
step: 380, loss: 0.012407103553414345
step: 390, loss: 0.0005824592662975192
step: 400, loss: 0.0004472040745895356
step: 410, loss: 0.00020952668273821473
step: 420, loss: 0.002436363138258457
step: 430, loss: 4.305763650336303e-05
step: 440, loss: 7.41363110137172e-05
step: 450, loss: 0.0002551399520598352
step: 460, loss: 0.00026024793623946607
step: 470, loss: 4.513749445322901e-05
step: 480, loss: 5.004960621590726e-05
step: 490, loss: 0.06070207059383392
step: 500, loss: 0.0009104131604544818
step: 510, loss: 0.00792752020061016
step: 520, loss: 0.0009310034802183509
step: 530, loss: 0.00022798690770287067
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9353932584269663, f1=0.9368616527390901,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037481472827494144
step: 10, loss: 0.0005174007965251803
step: 20, loss: 3.076610664720647e-05
step: 30, loss: 0.0002917931124102324
step: 40, loss: 0.08686114847660065
step: 50, loss: 0.00014891009777784348
step: 60, loss: 9.935843991115689e-05
step: 70, loss: 0.0010907617397606373

```

step: 80, loss: 0.006090895272791386  
step: 90, loss: 0.001150559401139617  
step: 100, loss: 2.8575805117725395e-05  
step: 110, loss: 0.003780636703595519  
step: 120, loss: 0.0014086398296058178  
step: 130, loss: 0.0002072201605187729  
step: 140, loss: 7.84313160693273e-05  
step: 150, loss: 8.510297629982233e-05  
step: 160, loss: 0.00011546877067303285  
step: 170, loss: 0.0006420233985409141  
step: 180, loss: 0.002502923598513007  
step: 190, loss: 0.0011029411107301712  
step: 200, loss: 0.0006675887852907181  
step: 210, loss: 7.581113459309563e-05  
step: 220, loss: 5.1900082326028496e-05  
step: 230, loss: 0.0010034902952611446  
step: 240, loss: 4.249278936185874e-05  
step: 250, loss: 0.02442183904349804  
step: 260, loss: 2.8181024390505627e-05  
step: 270, loss: 0.0011515652295202017  
step: 280, loss: 0.00016878294991329312  
step: 290, loss: 0.00628088740631938  
step: 300, loss: 0.011654986999928951  
step: 310, loss: 0.008312138728797436  
step: 320, loss: 0.00019042064377572387  
step: 330, loss: 0.0017769518308341503  
step: 340, loss: 9.529902308713645e-05  
step: 350, loss: 9.562970808474347e-05  
step: 360, loss: 6.356945232255384e-05  
step: 370, loss: 2.393449904047884e-05  
step: 380, loss: 5.274218347040005e-05  
step: 390, loss: 0.00025759206619113684  
step: 400, loss: 2.5644190827733837e-05  
step: 410, loss: 0.000252073397859931  
step: 420, loss: 6.296154606388882e-05  
step: 430, loss: 0.0001002452481770888  
step: 440, loss: 0.00013596804637927562  
step: 450, loss: 3.0594976124120876e-05  
step: 460, loss: 0.0127561679109931  
step: 470, loss: 0.020736336708068848  
step: 480, loss: 0.00017716982983984053  
step: 490, loss: 0.010527453385293484  
step: 500, loss: 0.0003708407748490572  
step: 510, loss: 2.761860378086567e-05  
step: 520, loss: 2.4914170353440568e-05  
step: 530, loss: 2.872838194889482e-05  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9395348837209302, f1=0.9317343173431735,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.01935127330944e-05
step: 10, loss: 3.213725358364172e-05
step: 20, loss: 2.746943755482789e-05
step: 30, loss: 3.673378159874119e-05
step: 40, loss: 0.0005291124107316136
step: 50, loss: 0.0010724639287218451
step: 60, loss: 0.0001647021563258022
step: 70, loss: 6.284029223024845e-05
step: 80, loss: 0.0003983147325925529
step: 90, loss: 4.601868931786157e-05
step: 100, loss: 2.906759618781507e-05
step: 110, loss: 2.8165766707388684e-05
step: 120, loss: 7.471739081665874e-05
step: 130, loss: 3.3738440833985806e-05
step: 140, loss: 3.6543635360430926e-05
step: 150, loss: 9.789947944227606e-05
step: 160, loss: 0.00035223711165599525
step: 170, loss: 0.005094393156468868
step: 180, loss: 3.1477549782721326e-05
step: 190, loss: 0.000200796450371854
step: 200, loss: 2.748820224951487e-05
step: 210, loss: 5.7225319324061275e-05
step: 220, loss: 3.482966712908819e-05
step: 230, loss: 0.00010598121298244223
step: 240, loss: 3.1700659747002646e-05
step: 250, loss: 3.2091556931845844e-05
step: 260, loss: 0.003703925060108304
step: 270, loss: 6.00519088038709e-05
step: 280, loss: 7.728017953922972e-05
step: 290, loss: 0.0006040938897058368
step: 300, loss: 0.00014690660464111716
step: 310, loss: 7.50499966670759e-05
step: 320, loss: 3.138805550406687e-05
step: 330, loss: 3.0400886316783726e-05
step: 340, loss: 0.0003362458373885602
step: 350, loss: 0.0009228195413015783
step: 360, loss: 0.0005869675660505891
step: 370, loss: 2.4567691070842557e-05
step: 380, loss: 0.00014762468344997615
step: 390, loss: 3.756694059120491e-05
step: 400, loss: 2.8906863008160144e-05
step: 410, loss: 0.00043002949678339064
```

```

step: 420, loss: 0.000699893687851727
step: 430, loss: 6.960146856727079e-05
step: 440, loss: 0.00011960169649682939
step: 450, loss: 8.197785064112395e-05
step: 460, loss: 0.0001449951814720407
step: 470, loss: 0.00011621011071838439
step: 480, loss: 4.434383299667388e-05
step: 490, loss: 0.0001324941113125533
step: 500, loss: 0.0011164085008203983
step: 510, loss: 3.3872776839416474e-05
step: 520, loss: 0.00048201423487626016
step: 530, loss: 0.00022417769650928676
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9341983317886932, f1=0.9326568265682658,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.223610348068178e-05
step: 10, loss: 0.00017013309116009623
step: 20, loss: 0.0003172433061990887
step: 30, loss: 0.0030385726131498814
step: 40, loss: 2.651585418789182e-05
step: 50, loss: 4.0351565985474735e-05
step: 60, loss: 6.592119461856782e-05
step: 70, loss: 5.4219530284171924e-05
step: 80, loss: 2.277215571666602e-05
step: 90, loss: 0.00025816154084168375
step: 100, loss: 4.4333210098557174e-05
step: 110, loss: 9.03427935554646e-05
step: 120, loss: 0.00010061737702926621
step: 130, loss: 0.0012830401537939906
step: 140, loss: 0.00019158209033776075
step: 150, loss: 4.2220122850267217e-05
step: 160, loss: 3.6797093343921006e-05
step: 170, loss: 3.588016625144519e-05
step: 180, loss: 3.585009108064696e-05
step: 190, loss: 4.6331537305377424e-05
step: 200, loss: 0.002137852134183049
step: 210, loss: 0.0007801845786161721
step: 220, loss: 3.810738417087123e-05
step: 230, loss: 0.000776775530539453
step: 240, loss: 4.76561181130819e-05
step: 250, loss: 0.05772492289543152
step: 260, loss: 2.6764220820041373e-05
step: 270, loss: 5.698503082385287e-05

```

```

step: 280, loss: 2.3642913220101036e-05
step: 290, loss: 3.863218807964586e-05
step: 300, loss: 0.00012342477566562593
step: 310, loss: 4.348381844465621e-05
step: 320, loss: 0.00016408767260145396
step: 330, loss: 4.942158193443902e-05
step: 340, loss: 7.95533342170529e-05
step: 350, loss: 0.00027496038819663227
step: 360, loss: 2.0782801584573463e-05
step: 370, loss: 3.846345134661533e-05
step: 380, loss: 0.0012786044972017407
step: 390, loss: 2.5498971808701754e-05
step: 400, loss: 2.5520475901430473e-05
step: 410, loss: 6.103435953264125e-05
step: 420, loss: 2.355803007958457e-05
step: 430, loss: 8.264194912044331e-05
step: 440, loss: 4.892770084552467e-05
step: 450, loss: 5.152769517735578e-05
step: 460, loss: 1.8227512555313297e-05
step: 470, loss: 0.00018721465312410146
step: 480, loss: 0.00010830132669070736
step: 490, loss: 3.910786836058833e-05
step: 500, loss: 6.712674075970426e-05
step: 510, loss: 0.00034435215638950467
step: 520, loss: 1.8693137462832965e-05
step: 530, loss: 2.392690657870844e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9386416861826699, f1=0.9329608938547487,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.765178891830146e-05
step: 10, loss: 2.9208829801063985e-05
step: 20, loss: 2.07419871003367e-05
step: 30, loss: 3.3939562854357064e-05
step: 40, loss: 2.221710565208923e-05
step: 50, loss: 4.6437075070571154e-05
step: 60, loss: 3.2225991162704304e-05
step: 70, loss: 6.333242345135659e-05
step: 80, loss: 5.701020563719794e-05
step: 90, loss: 1.9706349121406674e-05
step: 100, loss: 6.609566480619833e-05
step: 110, loss: 2.5263741918024607e-05
step: 120, loss: 1.9460108887869865e-05
step: 130, loss: 2.020180363615509e-05

```



```

step: 140, loss: 5.622411845251918e-05
step: 150, loss: 4.827448719879612e-05
step: 160, loss: 0.0009381304844282568
step: 170, loss: 6.063884575269185e-05
step: 180, loss: 3.186847607139498e-05
step: 190, loss: 6.893956015119329e-05
step: 200, loss: 3.198064223397523e-05
step: 210, loss: 4.572166290017776e-05
step: 220, loss: 4.163375706411898e-05
step: 230, loss: 0.0003567859821487218
step: 240, loss: 2.4414915969828144e-05
step: 250, loss: 3.407739131944254e-05
step: 260, loss: 1.7221665984834544e-05
step: 270, loss: 0.002790075959637761
step: 280, loss: 1.829460234148428e-05
step: 290, loss: 0.0003205359389539808
step: 300, loss: 2.100268284266349e-05
step: 310, loss: 3.200787978130393e-05
step: 320, loss: 2.56517123489175e-05
step: 330, loss: 6.26281471340917e-05
step: 340, loss: 3.334360371809453e-05
step: 350, loss: 0.00075416179606691
step: 360, loss: 0.0006096892757341266
step: 370, loss: 2.510735794203356e-05
step: 380, loss: 2.411627610854339e-05
step: 390, loss: 5.07957192894537e-05
step: 400, loss: 7.02582547091879e-05
step: 410, loss: 2.1252260921755806e-05
step: 420, loss: 1.9669121684273705e-05
step: 430, loss: 3.128267780994065e-05
step: 440, loss: 1.4424154869630001e-05
step: 450, loss: 3.510059468680993e-05
step: 460, loss: 0.00029718762380070984
step: 470, loss: 2.3009950382402167e-05
step: 480, loss: 2.121511170116719e-05
step: 490, loss: 2.447776023473125e-05
step: 500, loss: 4.888290277449414e-05
step: 510, loss: 3.09184797515627e-05
step: 520, loss: 6.040948937879875e-05
step: 530, loss: 2.51786332228221e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9398704902867714, f1=0.9348623853211009,
best_f1=0.9352319706017456
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 3.077260771533474e-05  
step: 10, loss: 5.1633152907015756e-05  
step: 20, loss: 2.0801613572984934e-05  
step: 30, loss: 0.00015340597019530833  
step: 40, loss: 2.5498089598841034e-05  
step: 50, loss: 2.4551951355533674e-05  
step: 60, loss: 4.61889649159275e-05  
step: 70, loss: 2.4031129214563407e-05  
step: 80, loss: 3.0207351301214658e-05  
step: 90, loss: 2.040285471593961e-05  
step: 100, loss: 2.4701283109607175e-05  
step: 110, loss: 0.0006685082335025072  
step: 120, loss: 0.0009770517935976386  
step: 130, loss: 2.3710454115644097e-05  
step: 140, loss: 1.7694741472951137e-05  
step: 150, loss: 5.317513569025323e-05  
step: 160, loss: 0.0007944712415337563  
step: 170, loss: 0.00024048591149039567  
step: 180, loss: 2.8764725357177667e-05  
step: 190, loss: 0.00011062823614338413  
step: 200, loss: 2.510749618522823e-05  
step: 210, loss: 1.770587005012203e-05  
step: 220, loss: 1.4167080735205673e-05  
step: 230, loss: 0.000767433550208807  
step: 240, loss: 2.8140044378233142e-05  
step: 250, loss: 6.106997898314148e-05  
step: 260, loss: 0.00021522186580114067  
step: 270, loss: 0.00016532160225324333  
step: 280, loss: 5.1480568799888715e-05  
step: 290, loss: 0.00016262619465123862  
step: 300, loss: 0.00020154779485892504  
step: 310, loss: 3.518134690239094e-05  
step: 320, loss: 4.963595711160451e-05  
step: 330, loss: 0.005565940868109465  
step: 340, loss: 4.9129186663776636e-05  
step: 350, loss: 2.5826491764746606e-05  
step: 360, loss: 2.9483655453077517e-05  
step: 370, loss: 1.853666799433995e-05  
step: 380, loss: 2.3170347049017437e-05  
step: 390, loss: 1.835405419114977e-05  
step: 400, loss: 1.988882286241278e-05  
step: 410, loss: 4.178160088486038e-05  
step: 420, loss: 2.6560435799183324e-05  
step: 430, loss: 2.5185923732351512e-05  
step: 440, loss: 3.6803674447583035e-05  
step: 450, loss: 3.8672784285154194e-05  
step: 460, loss: 6.193593435455114e-05  
step: 470, loss: 4.963980245520361e-05

```

step: 480, loss: 2.0141958884778433e-05
step: 490, loss: 5.916256850468926e-05
step: 500, loss: 0.0004038949264213443
step: 510, loss: 0.0003948915400542319
step: 520, loss: 3.8113743357826024e-05
step: 530, loss: 3.124486465821974e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9377323420074348, f1=0.9349930843706776,
best_f1=0.9352319706017456
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:24, 236.95it/s]
load_f1 = 0.9418334108887855
real_f1 = 0.9408476944573824
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 247.83it/s]

```

### 3 DK

```
[ ]: # Para ejecutar dk necesita
      #!pip install -r requirements.txt
      !python -m spacy download en_core_web_lg
```

Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/public/simple/>

Collecting en\_core\_web\_lg==2.2.5

Downloading [https://github.com/explosion/spacy-models/releases/download/en\\_core\\_web\\_lg-2.2.5/en\\_core\\_web\\_lg-2.2.5.tar.gz](https://github.com/explosion/spacy-models/releases/download/en_core_web_lg-2.2.5/en_core_web_lg-2.2.5.tar.gz) (827.9 MB)

| | 827.9 MB 1.2 MB/s

Requirement already satisfied: spacy>=2.2.2 in  
/usr/local/lib/python3.7/dist-packages (from en\_core\_web\_lg==2.2.5) (2.2.3)

Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (0.10.1)

Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (1.0.0)

Requirement already satisfied: srsly<1.1.0,>=0.1.0 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (1.0.5)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (2.0.6)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (2.23.0)

Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-  
packages (from spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (1.19.2)

Requirement already satisfied: thinc<7.4.0,>=7.3.0 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (7.3.1)

Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-  
packages (from spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (57.4.0)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (1.0.7)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (3.0.6)

Requirement already satisfied: plac<1.2.0,>=0.9.6 in  
/usr/local/lib/python3.7/dist-packages (from  
spacy>=2.2.2->en\_core\_web\_lg==2.2.5) (1.1.3)

Requirement already satisfied: blis<0.5.0,>=0.4.0 in

```

/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (0.4.1)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.12.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5)
(3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5)
(4.1.1)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5)
(2.10)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (1.25.4)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (2022.6.15)
Requirement already satisfied: tqdm<5.0.0,>=4.10.0 in
/usr/local/lib/python3.7/dist-packages (from
thinc<7.4.0,>=7.3.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.41.0)
Building wheels for collected packages: en-core-web-lg
  Building wheel for en-core-web-lg (setup.py) ... done
  Created wheel for en-core-web-lg: filename=en_core_web_lg-2.2.5-py3-none-
any.whl size=829180942
sha256=0bfe3aec166f2443541c4837c371e24ed9e1bea7c3af27e8beab400de23a0751
  Stored in directory: /tmp/pip-ephem-wheel-cache-
cobnmchx/wheels/11/95/ba/2c36cc368c0bd339b44a791c2c1881a1fb714b78c29a4cb8f5
Successfully built en-core-web-lg
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-2.2.5
  Download and installation successful
You can now load the model via spacy.load('en_core_web_lg')

```

## 3.1 DK STRUCTURED

### 3.1.1 Beer - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Beer \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8690328598022461
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.2857142857142857, f1=0.2692307692307693,
best_f1=0.2692307692307693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37182462215423584
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.2857142857142857, f1=0.2692307692307693,
best_f1=0.2692307692307693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.324365496635437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.4090909090909091, f1=0.3333333333333326,
best_f1=0.3333333333333326
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.35414010286331177
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.5098039215686275, f1=0.32876712328767127,
best_f1=0.32876712328767127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2617083489894867
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.4680851063829786, f1=0.3870967741935483,
best_f1=0.32876712328767127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.27800503373146057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.4444444444444444, f1=0.38596491228070184,
best_f1=0.32876712328767127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19567441940307617
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.5, f1=0.38596491228070184, best_f1=0.32876712328767127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3075484335422516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.5555555555555556, f1=0.4285714285714286,
best_f1=0.4285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15831321477890015
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.5238095238095237, f1=0.3773584905660377,
best_f1=0.4285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22456605732440948
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.5294117647058824, f1=0.4444444444444444,
best_f1=0.4285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19968098402023315
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.5454545454545454, f1=0.45, best_f1=0.4285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.216544508934021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.5806451612903226, f1=0.4878048780487805,
best_f1=0.4878048780487805
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15921273827552795
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 13: dev_f1=0.5833333333333334, f1=0.48484848484848486,
best_f1=0.48484848484848486
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18639668822288513
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.5833333333333334, f1=0.48484848484848486,
best_f1=0.48484848484848486
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.30583655834198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.5833333333333334, f1=0.48484848484848486,
best_f1=0.48484848484848486
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
91it [00:00, 125264.74it/s]
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
load_f1 = 0.6153846153846153
real_f1 = 0.5517241379310344
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 426.03it/s]

```

### 3.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/DBLP-ACM \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --dk general \
    --dk product \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt

```

```
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
```

```

UserWarning)
step: 0, loss: 0.8065325021743774
step: 10, loss: 0.46080705523490906
step: 20, loss: 0.5875399708747864
step: 30, loss: 0.4420463740825653
step: 40, loss: 0.22823096811771393
step: 50, loss: 0.06340950727462769
step: 60, loss: 0.14764507114887238
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.15206646919250488
step: 80, loss: 0.15812434256076813
step: 90, loss: 0.03801741451025009
step: 100, loss: 0.0196636114269495
step: 110, loss: 0.020550429821014404
step: 120, loss: 0.010052488185465336
step: 130, loss: 0.013836403377354145
step: 140, loss: 0.10879622399806976
step: 150, loss: 0.09302080422639847
step: 160, loss: 0.296448677778244
step: 170, loss: 0.0032508994918316603
step: 180, loss: 0.010421440936625004
step: 190, loss: 0.010314390063285828
step: 200, loss: 0.019562959671020508
step: 210, loss: 0.00354587659239769
step: 220, loss: 0.005997044499963522
step: 230, loss: 0.020672300830483437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9755011135857461, f1=0.9775280898876404,
best_f1=0.9775280898876404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07044924050569534
step: 10, loss: 0.0011113674845546484
step: 20, loss: 0.011491873301565647
step: 30, loss: 0.01932751201093197
step: 40, loss: 0.003824308980256319
step: 50, loss: 0.1379307508468628
step: 60, loss: 0.0051027811132371426
step: 70, loss: 0.26337069272994995
step: 80, loss: 0.0731077492237091
step: 90, loss: 0.010294361039996147

```

```

step: 100, loss: 0.06298366189002991
step: 110, loss: 0.10095436125993729
step: 120, loss: 0.0036026732996106148
step: 130, loss: 0.00244580558501184
step: 140, loss: 0.09281895309686661
step: 150, loss: 0.0064107500948011875
step: 160, loss: 0.006452149711549282
step: 170, loss: 0.08707758039236069
step: 180, loss: 0.0029383539222180843
step: 190, loss: 0.07752759009599686
step: 200, loss: 0.0077403695322573185
step: 210, loss: 0.09984627366065979
step: 220, loss: 0.0013842190383002162
step: 230, loss: 0.20023643970489502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9841628959276018, f1=0.9806157354618015,
best_f1=0.9806157354618015
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010176498908549547
step: 10, loss: 0.00791053008288145
step: 20, loss: 0.007573496550321579
step: 30, loss: 0.03276742249727249
step: 40, loss: 0.020292285829782486
step: 50, loss: 0.00228065880946815
step: 60, loss: 0.008080938830971718
step: 70, loss: 0.0005387632409110665
step: 80, loss: 0.05537407472729683
step: 90, loss: 0.0006158885662443936
step: 100, loss: 0.0016647066222503781
step: 110, loss: 0.004855509847402573
step: 120, loss: 0.0022160836961120367
step: 130, loss: 0.010048536583781242
step: 140, loss: 0.0018334604101255536
step: 150, loss: 0.0024225586093962193
step: 160, loss: 0.004362208768725395
step: 170, loss: 0.005996945779770613
step: 180, loss: 0.0022790690418332815
step: 190, loss: 0.024906640872359276
step: 200, loss: 0.0025601612869650126
step: 210, loss: 0.043311357498168945

```

```

step: 220, loss: 0.0040335217490792274
step: 230, loss: 0.036356136202812195
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.987598647125141, f1=0.9830124575311437,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022003263235092163
step: 10, loss: 0.001455108285881579
step: 20, loss: 0.000518179964274168
step: 30, loss: 0.0004809790407307446
step: 40, loss: 0.026034867390990257
step: 50, loss: 0.00855322740972042
step: 60, loss: 0.0007688601035624743
step: 70, loss: 0.11266733705997467
step: 80, loss: 0.0535941906273365
step: 90, loss: 0.060065481811761856
step: 100, loss: 0.0010225982405245304
step: 110, loss: 0.0027303036767989397
step: 120, loss: 0.06649945676326752
step: 130, loss: 0.0019989360589534044
step: 140, loss: 0.001487201894633472
step: 150, loss: 0.001229261513799429
step: 160, loss: 0.0005671786493621767
step: 170, loss: 0.004284643102437258
step: 180, loss: 0.08179736137390137
step: 190, loss: 0.007773961406201124
step: 200, loss: 0.000603408319875598
step: 210, loss: 0.00025636982172727585
step: 220, loss: 0.0003288066072855145
step: 230, loss: 0.000567179755307734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.987736900780379, f1=0.9799554565701558,
best_f1=0.9799554565701558
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00040394204552285373
step: 10, loss: 0.022639373317360878
step: 20, loss: 0.00024679579655639827
step: 30, loss: 0.00022144499234855175
step: 40, loss: 0.0008613818790763617
step: 50, loss: 0.00027973350370302796
step: 60, loss: 0.00048274907749146223
step: 70, loss: 0.00011740771878976375
step: 80, loss: 0.0006154430448077619
step: 90, loss: 0.001322026364505291
step: 100, loss: 0.0077787404879927635
step: 110, loss: 0.000327479065163061
step: 120, loss: 0.07489006221294403
step: 130, loss: 0.012813098728656769
step: 140, loss: 0.0003122867492493242
step: 150, loss: 0.0014016850618645549
step: 160, loss: 0.007844977080821991
step: 170, loss: 0.0031089945696294308
step: 180, loss: 0.0005715179722756147
step: 190, loss: 0.0008941615233197808
step: 200, loss: 0.0015930254012346268
step: 210, loss: 0.0014820029027760029
step: 220, loss: 0.0007694008527323604
step: 230, loss: 0.002765989862382412
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9820627802690582, f1=0.978675645342312,
best_f1=0.9799554565701558
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038803627248853445
step: 10, loss: 0.004332021810114384
step: 20, loss: 0.001449780073016882
step: 30, loss: 0.0007009501568973064
step: 40, loss: 0.0010596816428005695
step: 50, loss: 0.023804068565368652
step: 60, loss: 0.036639418452978134
step: 70, loss: 0.0011686322977766395
step: 80, loss: 0.0066357264295220375
step: 90, loss: 0.017618538811802864
step: 100, loss: 0.00044603910646401346
step: 110, loss: 0.12820737063884735
step: 120, loss: 0.0007838260498829186
step: 130, loss: 0.003135545877739787

```



```

step: 140, loss: 0.0009757544612511992
step: 150, loss: 0.0006912047974765301
step: 160, loss: 0.0003144039656035602
step: 170, loss: 0.0015857785474509
step: 180, loss: 0.0010484347585588694
step: 190, loss: 0.0006971368566155434
step: 200, loss: 0.0006533534033223987
step: 210, loss: 0.00030220442567951977
step: 220, loss: 0.0020611402578651905
step: 230, loss: 0.007940290495753288
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9853107344632768, f1=0.9818594104308391,
best_f1=0.9799554565701558
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12395728379487991
step: 10, loss: 0.0003939216258004308
step: 20, loss: 0.0016492047579959035
step: 30, loss: 0.00044611046905629337
step: 40, loss: 0.014942754060029984
step: 50, loss: 0.00013266372843645513
step: 60, loss: 0.005172884091734886
step: 70, loss: 0.00044028673437424004
step: 80, loss: 0.00034910591784864664
step: 90, loss: 0.009223176166415215
step: 100, loss: 0.005878982599824667
step: 110, loss: 0.03184189647436142
step: 120, loss: 0.000735304900445044
step: 130, loss: 0.00024156882136594504
step: 140, loss: 0.00011712244304362684
step: 150, loss: 0.00016037739987950772
step: 160, loss: 0.0001866497623268515
step: 170, loss: 0.0002434324414934963
step: 180, loss: 0.000560083135496825
step: 190, loss: 0.0005007667932659388
step: 200, loss: 0.006038322579115629
step: 210, loss: 0.0053567178547382355
step: 220, loss: 0.0013631165493279696
step: 230, loss: 0.01435706578195095
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9898762654668166, f1=0.983050847457627,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01974157989025116
step: 10, loss: 0.06966972351074219
step: 20, loss: 0.000410504057072103
step: 30, loss: 0.0002969328488688916
step: 40, loss: 0.00016528840933460742
step: 50, loss: 0.00021830025070812553
step: 60, loss: 0.06204420328140259
step: 70, loss: 0.0001466220273869112
step: 80, loss: 0.0003136764862574637
step: 90, loss: 0.00026171517674811184
step: 100, loss: 0.00019147101556882262
step: 110, loss: 0.00038501809467561543
step: 120, loss: 0.000371218224262506
step: 130, loss: 0.0005033675115555525
step: 140, loss: 0.00014226064376998693
step: 150, loss: 0.00011118633119622245
step: 160, loss: 0.0004533107567112893
step: 170, loss: 0.0012518049916252494
step: 180, loss: 0.0009017399861477315
step: 190, loss: 0.0010505589889362454
step: 200, loss: 0.014516360126435757
step: 210, loss: 8.608162897871807e-05
step: 220, loss: 0.00011125497258035466
step: 230, loss: 0.08882750570774078
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9863945578231292, f1=0.9828962371721778,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00027119406149722636
step: 10, loss: 0.0008800671203061938
step: 20, loss: 0.0008409327710978687
step: 30, loss: 0.00037886278005316854
step: 40, loss: 7.406784425256774e-05
step: 50, loss: 0.00019854839774779975
step: 60, loss: 0.0003640788490884006
step: 70, loss: 0.0002718543109949678
step: 80, loss: 0.05081091448664665
step: 90, loss: 0.003257693024352193

```

```

step: 100, loss: 0.0004921816871501505
step: 110, loss: 0.00017366043175570667
step: 120, loss: 0.03741693124175072
step: 130, loss: 9.46942891459912e-05
step: 140, loss: 0.00012212505680508912
step: 150, loss: 6.223338277777657e-05
step: 160, loss: 0.00016202300321310759
step: 170, loss: 6.140166078694165e-05
step: 180, loss: 0.00011889974121004343
step: 190, loss: 5.606895865639672e-05
step: 200, loss: 0.000137993018142879
step: 210, loss: 9.505557682132348e-05
step: 220, loss: 0.07467282563447952
step: 230, loss: 0.0010596509091556072
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9852774631936579, f1=0.9828962371721778,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.556811710353941e-05
step: 10, loss: 8.553686348022893e-05
step: 20, loss: 7.96115054981783e-05
step: 30, loss: 0.00010256358655169606
step: 40, loss: 6.166879029478878e-05
step: 50, loss: 0.00015937922580633312
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.01630130410194397
step: 70, loss: 0.00025658230879344046
step: 80, loss: 0.00012334261555224657
step: 90, loss: 8.225666533689946e-05
step: 100, loss: 9.249962749890983e-05
step: 110, loss: 0.00020858217612840235
step: 120, loss: 0.0215322133153677
step: 130, loss: 8.824427641229704e-05
step: 140, loss: 0.005793215706944466
step: 150, loss: 6.673984171357006e-05
step: 160, loss: 0.00016099034110084176
step: 170, loss: 7.181725959526375e-05
step: 180, loss: 0.00012587386299856007
step: 190, loss: 8.84741239133291e-05
step: 200, loss: 0.00012039187276968732
step: 210, loss: 7.349594670813531e-05
step: 220, loss: 0.00020798345212824643
step: 230, loss: 6.25001048319973e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9853438556933484, f1=0.9875424688561721,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.967306737555191e-05
step: 10, loss: 0.00018630377599038184
step: 20, loss: 4.3917745642829686e-05
step: 30, loss: 8.669094677316025e-05
step: 40, loss: 0.0005661796894855797
step: 50, loss: 0.00577185396105051
step: 60, loss: 7.30280444258824e-05
step: 70, loss: 0.00013670370390173048
step: 80, loss: 0.0005567229236476123
step: 90, loss: 5.603424142464064e-05
step: 100, loss: 8.594543760409579e-05
step: 110, loss: 0.0003555850125849247
step: 120, loss: 7.25911304471083e-05
step: 130, loss: 0.00015441601863130927
step: 140, loss: 0.00018925528274849057
step: 150, loss: 5.214245538809337e-05
step: 160, loss: 6.504965131171048e-05
step: 170, loss: 0.0029497917275875807
step: 180, loss: 0.00010853906132979318
step: 190, loss: 0.0001393209386151284
step: 200, loss: 6.98546355124563e-05
step: 210, loss: 5.836716809426434e-05
step: 220, loss: 7.524272223236039e-05
step: 230, loss: 0.027183709666132927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9818594104308391, f1=0.9817351598173515,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017565950111020356
step: 10, loss: 0.00016310843056999147
step: 20, loss: 6.281105015659705e-05
step: 30, loss: 0.00011553767399163917
step: 40, loss: 5.8949171943822876e-05
step: 50, loss: 5.1871949835913256e-05
step: 60, loss: 0.016889367252588272
step: 70, loss: 0.0002527358301449567
step: 80, loss: 0.00022273225476965308

```

```

step: 90, loss: 0.00015591106784995645
step: 100, loss: 6.301726534729823e-05
step: 110, loss: 0.00442528584972024
step: 120, loss: 0.0001567792205605656
step: 130, loss: 8.550698112230748e-05
step: 140, loss: 6.622354703722522e-05
step: 150, loss: 8.7300970335491e-05
step: 160, loss: 0.04265903681516647
step: 170, loss: 6.440166907850653e-05
step: 180, loss: 5.60973712708801e-05
step: 190, loss: 0.0002604296605568379
step: 200, loss: 0.007817014120519161
step: 210, loss: 6.516183202620596e-05
step: 220, loss: 0.00950019434094429
step: 230, loss: 9.703383693704382e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9841986455981941, f1=0.9841269841269841,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.723488281248137e-05
step: 10, loss: 0.00012677116319537163
step: 20, loss: 8.846117270877585e-05
step: 30, loss: 0.015244925394654274
step: 40, loss: 0.00013214952195994556
step: 50, loss: 0.0001022621727315709
step: 60, loss: 6.172944267746061e-05
step: 70, loss: 0.00012210830755066127
step: 80, loss: 4.819495006813668e-05
step: 90, loss: 5.797976700705476e-05
step: 100, loss: 9.09633599803783e-05
step: 110, loss: 7.924588862806559e-05
step: 120, loss: 9.252162271877751e-05
step: 130, loss: 0.0001243531733052805
step: 140, loss: 0.0002744028461165726
step: 150, loss: 0.024849338456988335
step: 160, loss: 5.642585892928764e-05
step: 170, loss: 7.16520007699728e-05
step: 180, loss: 0.00018539866141509265
step: 190, loss: 4.7266872570617124e-05
step: 200, loss: 0.0006999806500971317
step: 210, loss: 7.792809628881514e-05
step: 220, loss: 8.441787213087082e-05
step: 230, loss: 0.00021524348994717002
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9875424688561721, f1=0.9852774631936579,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001115501654567197
step: 10, loss: 0.0001340339076705277
step: 20, loss: 0.0007799921440891922
step: 30, loss: 9.081469033844769e-05
step: 40, loss: 5.1137620175722986e-05
step: 50, loss: 0.00010973178723361343
step: 60, loss: 0.0011377213522791862
step: 70, loss: 0.0022373106330633163
step: 80, loss: 0.00011407376587158069
step: 90, loss: 0.00021858842228539288
step: 100, loss: 0.008090058341622353
step: 110, loss: 0.00013500070781446993
step: 120, loss: 0.00011549380724318326
step: 130, loss: 9.361012052977458e-05
step: 140, loss: 0.0005461975815705955
step: 150, loss: 8.203677862184122e-05
step: 160, loss: 4.3422078306321055e-05
step: 170, loss: 8.5010877228342e-05
step: 180, loss: 0.00013941392535343766
step: 190, loss: 0.005987232550978661
step: 200, loss: 5.030661486671306e-05
step: 210, loss: 0.00951273925602436
step: 220, loss: 0.0001913503947434947
step: 230, loss: 2.704119287955109e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9875706214689265, f1=0.9841628959276018,
best_f1=0.983050847457627
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.3987900173524395e-05
step: 10, loss: 0.0001146428840002045
step: 20, loss: 0.00010556750203249976
step: 30, loss: 0.00016129481082316488
step: 40, loss: 0.00011093150533270091
step: 50, loss: 9.667142148828134e-05
step: 60, loss: 3.734741039806977e-05
step: 70, loss: 8.732645801501349e-05
step: 80, loss: 0.0013339589349925518

```

```

step: 90, loss: 4.250144775141962e-05
step: 100, loss: 0.002094193594530225
step: 110, loss: 8.554019586881623e-05
step: 120, loss: 0.00013489836419466883
step: 130, loss: 9.125618817051873e-05
step: 140, loss: 0.00010218022362096235
step: 150, loss: 0.00010446824308019131
step: 160, loss: 0.00017604767344892025
step: 170, loss: 5.7938570535043254e-05
step: 180, loss: 7.464188820449635e-05
step: 190, loss: 8.003209950402379e-05
step: 200, loss: 6.0707825468853116e-05
step: 210, loss: 0.0002683360071387142
step: 220, loss: 0.00019160444207955152
step: 230, loss: 0.0009068402578122914
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9875706214689265, f1=0.9841628959276018,
best_f1=0.983050847457627
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```

packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 330.43it/s]
load_f1 = 0.9863945578231292
real_f1 = 0.984090909090909
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:11, 390.42it/s]

```

### 3.1.3 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.



```

Defaults for this optimization level are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7894861102104187
step: 10, loss: 0.41306692361831665
step: 20, loss: 0.47739896178245544
step: 30, loss: 0.3918520510196686
step: 40, loss: 0.26209741830825806
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.19052600860595703
step: 60, loss: 0.12652850151062012
step: 70, loss: 0.0845339223742485
step: 80, loss: 0.11741384118795395
step: 90, loss: 0.18512475490570068
step: 100, loss: 0.25629740953445435
step: 110, loss: 0.09276416152715683
step: 120, loss: 0.05731482058763504

```

```

step: 130, loss: 0.039174243807792664
step: 140, loss: 0.20986106991767883
step: 150, loss: 0.09398902207612991
step: 160, loss: 0.09151109308004379
step: 170, loss: 0.19789110124111176
step: 180, loss: 0.1702239066362381
step: 190, loss: 0.03525468334555626
step: 200, loss: 0.1872216761112213
step: 210, loss: 0.128228098154068
step: 220, loss: 0.056654974818229675
step: 230, loss: 0.10538836568593979
step: 240, loss: 0.1867106556892395
step: 250, loss: 0.07703551650047302
step: 260, loss: 0.019805319607257843
step: 270, loss: 0.017326774075627327
step: 280, loss: 0.12097717821598053
step: 290, loss: 0.18069790303707123
step: 300, loss: 0.07645705342292786
step: 310, loss: 0.05684181675314903
step: 320, loss: 0.029026804491877556
step: 330, loss: 0.13897940516471863
step: 340, loss: 0.2165893018245697
step: 350, loss: 0.05472828075289726
step: 360, loss: 0.07334045320749283
step: 370, loss: 0.18720941245555878
step: 380, loss: 0.24689187109470367
step: 390, loss: 0.024618618190288544
step: 400, loss: 0.04661145806312561
step: 410, loss: 0.03179898485541344
step: 420, loss: 0.02159879170358181
step: 430, loss: 0.052623048424720764
step: 440, loss: 0.14897042512893677
step: 450, loss: 0.021599464118480682
step: 460, loss: 0.13230234384536743
step: 470, loss: 0.20367059111595154
step: 480, loss: 0.2733010947704315
step: 490, loss: 0.021149154752492905
step: 500, loss: 0.01881418190896511
step: 510, loss: 0.07819587737321854
step: 520, loss: 0.05084054917097092
step: 530, loss: 0.10030068457126617
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```
epoch 1: dev_f1=0.9262180974477957, f1=0.9274156264447527,  
best_f1=0.9274156264447527  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.0903659239411354  
step: 10, loss: 0.1269679218530655  
step: 20, loss: 0.10511299222707748  
step: 30, loss: 0.02989371493458748  
step: 40, loss: 0.007733039557933807  
step: 50, loss: 0.11156415194272995  
step: 60, loss: 0.12077581137418747  
step: 70, loss: 0.11581910401582718  
step: 80, loss: 0.005694198422133923  
step: 90, loss: 0.03226825222373009  
step: 100, loss: 0.24967089295387268  
step: 110, loss: 0.012461154721677303  
step: 120, loss: 0.14188919961452484  
step: 130, loss: 0.05869811773300171  
step: 140, loss: 0.01858341507613659  
step: 150, loss: 0.05210939422249794  
step: 160, loss: 0.046149492263793945  
step: 170, loss: 0.039468757808208466  
step: 180, loss: 0.009891683235764503  
step: 190, loss: 0.01727655529975891  
step: 200, loss: 0.02193969488143921  
step: 210, loss: 0.03301701322197914  
step: 220, loss: 0.15192942321300507  
step: 230, loss: 0.03173287957906723  
step: 240, loss: 0.1254691630601883  
step: 250, loss: 0.03670484200119972  
step: 260, loss: 0.008689646609127522  
step: 270, loss: 0.08995015174150467  
step: 280, loss: 0.3625902235507965  
step: 290, loss: 0.09953728318214417  
step: 300, loss: 0.024945564568042755  
step: 310, loss: 0.16253864765167236  
step: 320, loss: 0.1824425458908081  
step: 330, loss: 0.02419017069041729  
step: 340, loss: 0.01456264778971672  
step: 350, loss: 0.04471218213438988  
step: 360, loss: 0.04992321506142616  
step: 370, loss: 0.006696013268083334  
step: 380, loss: 0.13918140530586243  
step: 390, loss: 0.052489880472421646  
step: 400, loss: 0.02924177423119545  
step: 410, loss: 0.0021035876125097275  
step: 420, loss: 0.05333344265818596
```

```

step: 430, loss: 0.029502326622605324
step: 440, loss: 0.011410785838961601
step: 450, loss: 0.02322336472570896
step: 460, loss: 0.17630983889102936
step: 470, loss: 0.17923419177532196
step: 480, loss: 0.256330281496048
step: 490, loss: 0.04887102171778679
step: 500, loss: 0.03831451013684273
step: 510, loss: 0.07122049480676651
step: 520, loss: 0.09723746031522751
step: 530, loss: 0.11042246967554092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.933456561922366, f1=0.9344413665743306,
best_f1=0.9344413665743306
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04160642996430397
step: 10, loss: 0.13420869410037994
step: 20, loss: 0.18184630572795868
step: 30, loss: 0.16573192179203033
step: 40, loss: 0.0083708381280303
step: 50, loss: 0.008739386685192585
step: 60, loss: 0.0060993945226073265
step: 70, loss: 0.03283047676086426
step: 80, loss: 0.00804300419986248
step: 90, loss: 0.09388121962547302
step: 100, loss: 0.07249841839075089
step: 110, loss: 0.014520869590342045
step: 120, loss: 0.003911765292286873
step: 130, loss: 0.062458448112010956
step: 140, loss: 0.019660333171486855
step: 150, loss: 0.004915053024888039
step: 160, loss: 0.0022045939695090055
step: 170, loss: 0.003313911845907569
step: 180, loss: 0.015388129279017448
step: 190, loss: 0.0038530488964170218
step: 200, loss: 0.005730103235691786
step: 210, loss: 0.012490707449615002
step: 220, loss: 0.0892234593629837
step: 230, loss: 0.031223053112626076
step: 240, loss: 0.007270445581525564

```

```

step: 250, loss: 0.013278404250741005
step: 260, loss: 0.1020926684141159
step: 270, loss: 0.010563173331320286
step: 280, loss: 0.012620282359421253
step: 290, loss: 0.07289246469736099
step: 300, loss: 0.08520326018333435
step: 310, loss: 0.12887123227119446
step: 320, loss: 0.06162272021174431
step: 330, loss: 0.001991154393181205
step: 340, loss: 0.0041986326687037945
step: 350, loss: 0.026262924075126648
step: 360, loss: 0.0889570340514183
step: 370, loss: 0.008115525357425213
step: 380, loss: 0.02974480390548706
step: 390, loss: 0.06244603544473648
step: 400, loss: 0.015396473929286003
step: 410, loss: 0.04808356985449791
step: 420, loss: 0.04388401657342911
step: 430, loss: 0.0831981748342514
step: 440, loss: 0.09133048355579376
step: 450, loss: 0.033338554203510284
step: 460, loss: 0.15368881821632385
step: 470, loss: 0.005352542735636234
step: 480, loss: 0.022336415946483612
step: 490, loss: 0.003926469013094902
step: 500, loss: 0.02934473752975464
step: 510, loss: 0.007523622363805771
step: 520, loss: 0.00843545887619257
step: 530, loss: 0.039433252066373825
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9286047596826879, f1=0.9251510925151093,
best_f1=0.9344413665743306
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0048597133718431
step: 10, loss: 0.008687141351401806
step: 20, loss: 0.01303714420646429
step: 30, loss: 0.003875466762110591
step: 40, loss: 0.004797989968210459
step: 50, loss: 0.0035111764445900917
step: 60, loss: 0.0030609446112066507
step: 70, loss: 0.006636687088757753
step: 80, loss: 0.012488283216953278
step: 90, loss: 0.0029567719902843237
step: 100, loss: 0.007968869991600513

```

```
step: 110, loss: 0.0005475714569911361
step: 120, loss: 0.001300545409321785
step: 130, loss: 0.02338918298482895
step: 140, loss: 0.004247539676725864
step: 150, loss: 0.0007981591625139117
step: 160, loss: 0.0009319387609139085
step: 170, loss: 0.000647395325358957
step: 180, loss: 0.009798765182495117
step: 190, loss: 0.03199302777647972
step: 200, loss: 0.008034779690206051
step: 210, loss: 0.027457870543003082
step: 220, loss: 0.0020325786899775267
step: 230, loss: 0.1857810616493225
step: 240, loss: 0.01398381870239973
step: 250, loss: 0.0017071085749194026
step: 260, loss: 0.09068465977907181
step: 270, loss: 0.011422589421272278
step: 280, loss: 0.0005448011797852814
step: 290, loss: 0.06670217216014862
step: 300, loss: 0.06393782794475555
step: 310, loss: 0.005174410063773394
step: 320, loss: 0.10378530621528625
step: 330, loss: 0.028919091448187828
step: 340, loss: 0.016785426065325737
step: 350, loss: 0.008916333317756653
step: 360, loss: 0.01647110842168331
step: 370, loss: 0.11775683611631393
step: 380, loss: 0.01089646015316248
step: 390, loss: 0.04903604835271835
step: 400, loss: 0.020464349538087845
step: 410, loss: 0.05279488116502762
step: 420, loss: 0.004836540203541517
step: 430, loss: 0.010229027830064297
step: 440, loss: 0.07415526360273361
step: 450, loss: 0.019508739933371544
step: 460, loss: 0.00026916153728961945
step: 470, loss: 0.006399065721780062
step: 480, loss: 0.012929624877870083
step: 490, loss: 0.044474076479673386
step: 500, loss: 0.006834226194769144
step: 510, loss: 0.03076140210032463
step: 520, loss: 0.004399872850626707
step: 530, loss: 0.015870949253439903
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
```

```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9340710004610421, f1=0.9274156264447527,
best_f1=0.9274156264447527
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0054031554609537125
step: 10, loss: 0.0011891793692484498
step: 20, loss: 0.0014920863322913647
step: 30, loss: 0.0005754315643571317
step: 40, loss: 0.02644815668463707
step: 50, loss: 0.004163812845945358
step: 60, loss: 0.03295142576098442
step: 70, loss: 0.00024213569122366607
step: 80, loss: 0.0006967331864871085
step: 90, loss: 0.008990530855953693
step: 100, loss: 0.00013378103903960437
step: 110, loss: 0.029552491381764412
step: 120, loss: 0.0036121313460171223
step: 130, loss: 0.0015590160619467497
step: 140, loss: 0.0027815974317491055
step: 150, loss: 0.0002895992947742343
step: 160, loss: 0.05944870039820671
step: 170, loss: 0.009149040095508099
step: 180, loss: 0.0008393291500397027
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.0001427719253115356
step: 200, loss: 0.0003479689185041934
step: 210, loss: 0.0011119425762444735
step: 220, loss: 0.00095420639263466
step: 230, loss: 0.0009118287125602365
step: 240, loss: 0.004001120571047068
step: 250, loss: 0.0003172885626554489
step: 260, loss: 0.0002513204817660153
step: 270, loss: 0.005967874079942703
step: 280, loss: 0.026466483250260353
step: 290, loss: 0.0014072355115786195
step: 300, loss: 0.003551197238266468
step: 310, loss: 0.0003780537226703018
step: 320, loss: 0.007261510938405991
step: 330, loss: 0.0002852472825907171
step: 340, loss: 0.013813083060085773
step: 350, loss: 0.000782627728767693
step: 360, loss: 0.07286395877599716
step: 370, loss: 0.008044061250984669
step: 380, loss: 0.05413057282567024
step: 390, loss: 0.0067131295800209045

```

```

step: 400, loss: 0.06271934509277344
step: 410, loss: 0.00021863695292267948
step: 420, loss: 0.012566311284899712
step: 430, loss: 0.0018598750466480851
step: 440, loss: 0.005579219665378332
step: 450, loss: 0.001377689535729587
step: 460, loss: 0.0004222710558678955
step: 470, loss: 0.000662843813188374
step: 480, loss: 0.000575757585465908
step: 490, loss: 0.0056661274284124374
step: 500, loss: 0.013415651395916939
step: 510, loss: 0.00037619160139001906
step: 520, loss: 0.002023621927946806
step: 530, loss: 0.004604260437190533
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9351376574895007, f1=0.9343955014058105,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024794028140604496
step: 10, loss: 0.0007273147348314524
step: 20, loss: 0.00015099930169526488
step: 30, loss: 0.002118020551279187
step: 40, loss: 0.05194542929530144
step: 50, loss: 0.0003254490438848734
step: 60, loss: 0.008330244570970535
step: 70, loss: 0.0021264730021357536
step: 80, loss: 0.0006651939474977553
step: 90, loss: 0.018945815041661263
step: 100, loss: 0.009810621850192547
step: 110, loss: 0.037305232137441635
step: 120, loss: 0.0009340585675090551
step: 130, loss: 7.334620750043541e-05
step: 140, loss: 0.00044942705426365137
step: 150, loss: 0.00040251738391816616
step: 160, loss: 0.0005735133890993893
step: 170, loss: 0.0001091508602257818
step: 180, loss: 0.013806473463773727
step: 190, loss: 0.0001809409586712718
step: 200, loss: 6.923743785591796e-05
step: 210, loss: 9.334290371043608e-05

```



```

step: 220, loss: 0.0020737643353641033
step: 230, loss: 0.0001803369086701423
step: 240, loss: 0.000423512450652197
step: 250, loss: 0.04164360463619232
step: 260, loss: 0.00022084864031057805
step: 270, loss: 0.00017260879394598305
step: 280, loss: 0.00025774864479899406
step: 290, loss: 0.0012196011375635862
step: 300, loss: 0.00018742585962172598
step: 310, loss: 0.00018325625569559634
step: 320, loss: 0.026209663599729538
step: 330, loss: 0.002666980493813753
step: 340, loss: 0.04630313441157341
step: 350, loss: 0.026104046031832695
step: 360, loss: 0.0002433109621051699
step: 370, loss: 0.008438068442046642
step: 380, loss: 0.005982496775686741
step: 390, loss: 0.0194252822548151
step: 400, loss: 0.0009180358028970659
step: 410, loss: 0.0005423976108431816
step: 420, loss: 0.05205323547124863
step: 430, loss: 0.08480433374643326
step: 440, loss: 0.07888011634349823
step: 450, loss: 0.0006685908883810043
step: 460, loss: 0.03300594910979271
step: 470, loss: 0.12095259130001068
step: 480, loss: 0.005026671104133129
step: 490, loss: 0.0001741503510857001
step: 500, loss: 0.0019446746446192265
step: 510, loss: 0.00037378218257799745
step: 520, loss: 0.1076396033167839
step: 530, loss: 0.007122226059436798
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9299303944315545, f1=0.9254284390921722,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016459587961435318
step: 10, loss: 0.0041737062856554985
step: 20, loss: 0.001739308936521411
step: 30, loss: 0.016316937282681465
step: 40, loss: 7.581441604997963e-05
step: 50, loss: 0.0007153706392273307
step: 60, loss: 0.0004462914075702429
step: 70, loss: 0.0010314867831766605

```

```
step: 80, loss: 0.012616602703928947
step: 90, loss: 0.00015623871877323836
step: 100, loss: 0.018281590193510056
step: 110, loss: 0.007706168573349714
step: 120, loss: 0.0001341502065770328
step: 130, loss: 0.04172142595052719
step: 140, loss: 0.04224354401230812
step: 150, loss: 8.680191240273416e-05
step: 160, loss: 0.0006193149019964039
step: 180, loss: 0.00010218170791631564
step: 190, loss: 0.00012978802260477096
step: 200, loss: 0.02087467908859253
step: 210, loss: 0.0005435499479062855
step: 220, loss: 0.001448458177037537
step: 230, loss: 0.00012597157910931855
step: 240, loss: 0.005710485391318798
step: 250, loss: 0.00013337814016267657
step: 260, loss: 0.0022378633730113506
step: 270, loss: 0.00021068498608656228
step: 280, loss: 0.003154770238325
step: 290, loss: 0.007372613530606031
step: 300, loss: 0.0002350778377149254
step: 310, loss: 0.00011395323963370174
step: 320, loss: 0.024453742429614067
step: 330, loss: 0.00028717020177282393
step: 340, loss: 0.0005951342172920704
step: 350, loss: 0.003607708727940917
step: 360, loss: 0.004400161094963551
step: 370, loss: 0.0005187244969420135
step: 380, loss: 0.003557929303497076
step: 390, loss: 0.02516346052289009
step: 400, loss: 0.0001260720455320552
step: 410, loss: 0.0018366622971370816
step: 420, loss: 0.001898600603453815
step: 430, loss: 5.8831454225583e-05
step: 440, loss: 0.00020487811707425863
step: 450, loss: 0.00014765291416551918
step: 460, loss: 0.0012315531494095922
step: 470, loss: 0.001000349991954863
step: 480, loss: 7.896382157923654e-05
step: 490, loss: 0.0006014963146299124
step: 500, loss: 5.567324114963412e-05
step: 510, loss: 0.00044055478065274656
step: 520, loss: 0.0001281318545807153
step: 530, loss: 0.0008223415352404118
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9366494603472549, f1=0.929840972871843,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008294739527627826
step: 10, loss: 0.0017329048132523894
step: 20, loss: 0.004447738640010357
step: 30, loss: 0.0005547211039811373
step: 40, loss: 0.0001767021749401465
step: 50, loss: 8.428408909821883e-05
step: 60, loss: 3.8778372982051224e-05
step: 70, loss: 0.00021518352150451392
step: 80, loss: 8.626876660855487e-05
step: 90, loss: 6.970401591388509e-05
step: 100, loss: 0.0027893034275621176
step: 110, loss: 0.005439207423478365
step: 120, loss: 0.0007565103005617857
step: 130, loss: 0.0037642281968146563
step: 140, loss: 3.4416316339047626e-05
step: 150, loss: 8.274690480902791e-05
step: 160, loss: 0.00024302514793816954
step: 170, loss: 4.5169785153120756e-05
step: 180, loss: 7.315297261811793e-05
step: 190, loss: 0.0032056851778179407
step: 200, loss: 0.049054037779569626
step: 210, loss: 0.004238355439156294
step: 220, loss: 0.0008097966201603413
step: 230, loss: 0.00011403892131056637
step: 240, loss: 6.916160054970533e-05
step: 250, loss: 5.266103471512906e-05
step: 260, loss: 0.0019077848410233855
step: 270, loss: 5.859098746441305e-05
step: 280, loss: 0.00010462114732945338
step: 290, loss: 0.001079666893929243
step: 300, loss: 8.668539521750063e-05
step: 310, loss: 0.10924088209867477
step: 320, loss: 0.04648537188768387
step: 330, loss: 0.006927340757101774
step: 340, loss: 0.0011295800795778632
step: 350, loss: 0.0048707169480621815
step: 360, loss: 0.11607881635427475
step: 370, loss: 0.0015137207228690386
step: 380, loss: 0.00029342324705794454

```

```

step: 390, loss: 0.00031203654361888766
step: 400, loss: 0.021537797525525093
step: 410, loss: 0.00015048372733872384
step: 420, loss: 7.41107651265338e-05
step: 430, loss: 0.02090822532773018
step: 440, loss: 0.0011833319440484047
step: 450, loss: 0.0003663363168016076
step: 460, loss: 0.00028981882496736944
step: 470, loss: 0.020402293652296066
step: 480, loss: 0.00021690755966119468
step: 490, loss: 0.0016935761086642742
step: 500, loss: 7.852102862671018e-05
step: 510, loss: 0.000195432105101645
step: 520, loss: 0.00015866516332607716
step: 530, loss: 0.00011244796041864902
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9299303944315545, f1=0.9305108145421077,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001823760976549238
step: 10, loss: 0.0033073732629418373
step: 20, loss: 0.003482979256659746
step: 30, loss: 0.0005981970462016761
step: 40, loss: 0.011023021303117275
step: 50, loss: 0.0003416347608435899
step: 60, loss: 0.0005279277102090418
step: 70, loss: 0.09431981295347214
step: 80, loss: 0.00023848055570852011
step: 90, loss: 0.0004897167673334479
step: 100, loss: 0.0006230776198208332
step: 110, loss: 0.000835083716083318
step: 120, loss: 0.00015656383766327053
step: 130, loss: 0.00018976429419126362
step: 140, loss: 0.0005471569020301104
step: 150, loss: 0.0001675820822129026
step: 160, loss: 0.0007919674389995635
step: 170, loss: 0.002009175019338727
step: 180, loss: 8.300798799609765e-05
step: 190, loss: 0.0004983748658560216
step: 200, loss: 0.001204966683872044
step: 210, loss: 0.0004200713010504842
step: 220, loss: 0.0021403077989816666
step: 230, loss: 0.0005053639179095626
step: 240, loss: 0.0002456970978528261

```

```

step: 250, loss: 0.001066387863829732
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 260, loss: 0.00018811643531080335
step: 270, loss: 0.0005506266024895012
step: 280, loss: 0.00010887693497352302
step: 290, loss: 0.00012774611241184175
step: 300, loss: 0.0034466837532818317
step: 310, loss: 0.00013521626533474773
step: 320, loss: 0.021854382008314133
step: 330, loss: 6.961997132748365e-05
step: 340, loss: 0.004739142023026943
step: 350, loss: 8.600227010902017e-05
step: 360, loss: 0.014448355883359909
step: 370, loss: 0.0013108940329402685
step: 380, loss: 0.00015290947339963168
step: 390, loss: 0.001196559052914381
step: 400, loss: 0.0010175085626542568
step: 410, loss: 0.0002080301783280447
step: 420, loss: 0.0008230134262703359
step: 430, loss: 7.420634938171133e-05
step: 440, loss: 0.0010827853111550212
step: 450, loss: 0.000108142223325558
step: 460, loss: 0.0236731618642807
step: 470, loss: 7.591954636154696e-05
step: 480, loss: 8.327446266775951e-05
step: 490, loss: 0.00017531687626615167
step: 500, loss: 0.000547753821592778
step: 510, loss: 0.00024173397105187178
step: 520, loss: 9.682230302132666e-05
step: 530, loss: 0.0001748601789586246
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.936470588235294, f1=0.9305164319248826,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000305289780953899
step: 10, loss: 5.0803682825062424e-05
step: 20, loss: 0.0004884329391643405
step: 30, loss: 0.00013952968583907932
step: 40, loss: 0.003234265372157097
step: 50, loss: 5.570827852352522e-05
step: 60, loss: 8.797353802947327e-05
step: 70, loss: 0.0007205871515907347
step: 80, loss: 0.001880654483102262
step: 90, loss: 0.00011170515062985942

```

```
step: 100, loss: 9.676056652097031e-05
step: 110, loss: 0.005470690783113241
step: 120, loss: 0.00011429719597799703
step: 130, loss: 7.273171650012955e-05
step: 140, loss: 0.0003890578227583319
step: 150, loss: 8.294804865727201e-05
step: 160, loss: 0.00026159200933761895
step: 170, loss: 0.00034824557951651514
step: 180, loss: 0.012045549228787422
step: 190, loss: 0.0003780413535423577
step: 200, loss: 0.00013151438906788826
step: 210, loss: 8.256118599092588e-05
step: 220, loss: 0.0001649215555517003
step: 230, loss: 0.00018824319704435766
step: 240, loss: 6.440999277401716e-05
step: 250, loss: 0.00021177840244490653
step: 260, loss: 0.0016973139718174934
step: 270, loss: 0.00011806876864284277
step: 280, loss: 0.00040290827746503055
step: 290, loss: 4.2902080167550594e-05
step: 300, loss: 0.00376448268070817
step: 310, loss: 4.5806620619259775e-05
step: 320, loss: 0.0005073669017292559
step: 330, loss: 0.02483179047703743
step: 340, loss: 0.00019822771719191223
step: 350, loss: 0.00022812985116615891
step: 360, loss: 0.00012450246140360832
step: 370, loss: 0.0008206287166103721
step: 380, loss: 0.0001392084959661588
step: 390, loss: 0.0003637151967268437
step: 400, loss: 0.00014200617442838848
step: 410, loss: 0.00010470623965375125
step: 420, loss: 0.0013731635408475995
step: 430, loss: 9.523784683551639e-05
step: 440, loss: 0.00011837715283036232
step: 450, loss: 0.006593554746359587
step: 460, loss: 0.00038123465492390096
step: 470, loss: 0.00012055024853907526
step: 480, loss: 9.736007632454857e-05
step: 490, loss: 0.03085564449429512
step: 500, loss: 0.0028447131626307964
step: 510, loss: 0.0012006480246782303
step: 520, loss: 0.00011030197492800653
step: 530, loss: 4.175897629465908e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
```

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 10: dev\_f1=0.936953520478601, f1=0.9326614750343564,

best\_f1=0.9326614750343564

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.0005176355480216444

step: 10, loss: 0.0019037314923480153

step: 20, loss: 0.00012725035776384175

step: 30, loss: 0.0008205818012356758

step: 40, loss: 0.0001886155951069668

step: 50, loss: 0.0007404378848150373

step: 60, loss: 8.480651013087481e-05

step: 70, loss: 5.469785537570715e-05

step: 80, loss: 7.433111022692174e-05

step: 90, loss: 0.0005494972574524581

step: 100, loss: 0.0001707512856228277

step: 110, loss: 5.137879270478152e-05

step: 120, loss: 0.004313485696911812

step: 130, loss: 3.9940357964951545e-05

step: 140, loss: 3.390253186807968e-05

step: 150, loss: 5.5863238230813295e-05

step: 160, loss: 0.00013715388195123523

step: 170, loss: 0.0021374973002821207

step: 180, loss: 2.9682469175895676e-05

step: 190, loss: 0.0010253351647406816

step: 200, loss: 0.0007674349471926689

step: 210, loss: 0.0006592539139091969

step: 220, loss: 6.33465897408314e-05

step: 230, loss: 8.68673378136009e-05

step: 240, loss: 0.00048220480675809085

step: 250, loss: 7.54796274122782e-05

step: 260, loss: 4.071538205607794e-05

step: 270, loss: 0.028239436447620392

step: 280, loss: 0.0002438180526951328

step: 290, loss: 0.00019741224241442978

step: 300, loss: 0.007176430430263281

step: 310, loss: 0.0017036416102200747

step: 320, loss: 0.021774018183350563

step: 330, loss: 0.0028032807167619467

step: 340, loss: 0.020891083404421806

step: 350, loss: 0.04993688315153122

step: 360, loss: 0.0001485440443502739

step: 370, loss: 4.9212321755476296e-05

step: 380, loss: 0.04874829575419426

step: 390, loss: 0.0004910427378490567

```

step: 400, loss: 4.9618403863860294e-05
step: 410, loss: 0.00015862446161918342
step: 420, loss: 0.005520017351955175
step: 430, loss: 0.00010092376032844186
step: 440, loss: 8.715283183846623e-05
step: 450, loss: 0.00013021162885706872
step: 460, loss: 0.002562318928539753
step: 470, loss: 0.005816114135086536
step: 480, loss: 8.544504817109555e-05
step: 490, loss: 0.0005070645711384714
step: 500, loss: 0.0001927776465890929
step: 510, loss: 0.00011461544636404142
step: 520, loss: 8.532088395440951e-05
step: 530, loss: 3.6684974475065246e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.9395477618827872, f1=0.9283746556473829,
best_f1=0.9283746556473829
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017386752006132156
step: 10, loss: 6.726137507939711e-05
step: 20, loss: 5.167989365872927e-05
step: 30, loss: 4.770719533553347e-05
step: 40, loss: 9.692491585155949e-05
step: 50, loss: 0.0029206788167357445
step: 60, loss: 0.00010312676022294909
step: 70, loss: 0.0007100833463482559
step: 80, loss: 0.0017208572244271636
step: 90, loss: 5.613947723759338e-05
step: 100, loss: 0.00015450730279553682
step: 110, loss: 3.166773240081966e-05
step: 120, loss: 0.00018089477089233696
step: 130, loss: 7.963062671478838e-05
step: 140, loss: 4.939026985084638e-05
step: 150, loss: 0.00017956586088985205
step: 160, loss: 0.0001639464608160779
step: 170, loss: 0.0004785823111888021
step: 180, loss: 0.0008459854288958013
step: 190, loss: 2.398658762103878e-05
step: 200, loss: 0.0008125104941427708
step: 210, loss: 0.0029028167482465506

```



```

step: 220, loss: 2.4761520762695e-05
step: 230, loss: 0.0023744956124573946
step: 240, loss: 9.144842624664307e-05
step: 250, loss: 0.0005342781078070402
step: 260, loss: 4.568369695334695e-05
step: 270, loss: 5.1801674999296665e-05
step: 280, loss: 2.9566655939561315e-05
step: 290, loss: 0.023672815412282944
step: 300, loss: 5.194505138206296e-05
step: 310, loss: 5.014439739170484e-05
step: 320, loss: 3.299751551821828e-05
step: 330, loss: 5.4918269597692415e-05
step: 340, loss: 0.0001469933777116239
step: 350, loss: 0.023622969165444374
step: 360, loss: 0.0009525510831736028
step: 370, loss: 3.181301144650206e-05
step: 380, loss: 6.018561907694675e-05
step: 390, loss: 0.00028214568737894297
step: 400, loss: 3.6725919926539063e-05
step: 410, loss: 7.367664511548355e-05
step: 420, loss: 0.0010016672313213348
step: 430, loss: 0.015130081214010715
step: 440, loss: 4.649081529350951e-05
step: 450, loss: 9.981060429709032e-05
step: 460, loss: 8.764586527831852e-05
step: 470, loss: 0.0001235354138771072
step: 480, loss: 0.004629502538591623
step: 490, loss: 0.001078121829777956
step: 500, loss: 0.0016962018562480807
step: 510, loss: 4.646582965506241e-05
step: 520, loss: 0.023642471060156822
step: 530, loss: 5.387575947679579e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9382022471910113, f1=0.9314685314685315,
best_f1=0.9283746556473829
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.659269299940206e-05
step: 10, loss: 9.556628356222063e-05
step: 20, loss: 0.03494470566511154
step: 30, loss: 3.244996696594171e-05
step: 40, loss: 5.871350731467828e-05
step: 50, loss: 4.007088500657119e-05
step: 60, loss: 3.301949618617073e-05
step: 70, loss: 0.0008056406513787806

```

step: 80, loss: 2.6545425498625264e-05  
step: 90, loss: 9.377427340950817e-05  
step: 100, loss: 2.2045829609851353e-05  
step: 110, loss: 3.9951239159563556e-05  
step: 120, loss: 0.00019773424719460309  
step: 130, loss: 0.0026032226160168648  
step: 140, loss: 2.306648821104318e-05  
step: 150, loss: 3.9537168049719185e-05  
step: 160, loss: 2.6475017875782214e-05  
step: 170, loss: 3.8946680433582515e-05  
step: 180, loss: 2.307760587427765e-05  
step: 190, loss: 2.9130933398846537e-05  
step: 200, loss: 0.002027987502515316  
step: 210, loss: 0.007744513917714357  
step: 220, loss: 2.860856147890445e-05  
step: 230, loss: 2.9338867534534074e-05  
step: 240, loss: 2.284306174260564e-05  
step: 250, loss: 0.07623754441738129  
step: 260, loss: 8.248658559750766e-05  
step: 270, loss: 5.789374699816108e-05  
step: 280, loss: 2.2425601855502464e-05  
step: 290, loss: 2.4575239876867272e-05  
step: 300, loss: 3.240115984226577e-05  
step: 310, loss: 6.0351834690663964e-05  
step: 320, loss: 2.2001140678185038e-05  
step: 330, loss: 0.00030114816036075354  
step: 340, loss: 2.2593478206545115e-05  
step: 350, loss: 0.0003073012921959162  
step: 360, loss: 1.9434402929618955e-05  
step: 370, loss: 0.00010666460730135441  
step: 380, loss: 2.7726704502128996e-05  
step: 390, loss: 3.6549725336954e-05  
step: 400, loss: 3.1755364034324884e-05  
step: 410, loss: 3.516839205985889e-05  
step: 420, loss: 3.1000658054836094e-05  
step: 430, loss: 0.0006327539449557662  
step: 440, loss: 4.307613198761828e-05  
step: 450, loss: 2.8478936656028964e-05  
step: 460, loss: 2.5106628527282737e-05  
step: 470, loss: 0.02499573677778244  
step: 480, loss: 0.0016532752197235823  
step: 490, loss: 0.0001227332541020587  
step: 500, loss: 8.459383388981223e-05  
step: 510, loss: 3.337548332638107e-05  
step: 520, loss: 2.343147025385406e-05  
step: 530, loss: 5.127046460984275e-05  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9354536950420954, f1=0.9312267657992565,
best_f1=0.9283746556473829
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.373819633154199e-05
step: 10, loss: 0.00010076351463794708
step: 20, loss: 1.8182850908488035e-05
step: 30, loss: 0.0002535503008402884
step: 40, loss: 2.9033455575699918e-05
step: 50, loss: 2.730904088821262e-05
step: 60, loss: 2.457506889186334e-05
step: 70, loss: 2.4593233320047148e-05
step: 80, loss: 0.0015538973966613412
step: 90, loss: 2.1997097064740956e-05
step: 100, loss: 3.141787237836979e-05
step: 110, loss: 2.105473686242476e-05
step: 120, loss: 7.146518328227103e-05
step: 130, loss: 4.126280327909626e-05
step: 140, loss: 4.782461473951116e-05
step: 150, loss: 0.00019140700169373304
step: 160, loss: 0.0011970478808507323
step: 170, loss: 0.00013394020788837224
step: 180, loss: 0.0027832393534481525
step: 190, loss: 4.539919245871715e-05
step: 200, loss: 0.0016741097206249833
step: 210, loss: 0.00012166657688794658
step: 220, loss: 0.0002648335648700595
step: 230, loss: 0.0018005111487582326
step: 240, loss: 0.027377400547266006
step: 250, loss: 3.651290535344742e-05
step: 260, loss: 1.8618737158249132e-05
step: 270, loss: 0.003648460377007723
step: 280, loss: 3.6325440305517986e-05
step: 290, loss: 3.466703128651716e-05
step: 300, loss: 2.8467429729062133e-05
step: 310, loss: 2.5301285859313793e-05
step: 320, loss: 6.848773773526773e-05
step: 330, loss: 2.5769866624614224e-05
step: 340, loss: 3.534369534463622e-05
step: 350, loss: 0.0010752595262601972
step: 360, loss: 0.003533317241817713
step: 370, loss: 0.005880286451429129
step: 380, loss: 3.7292862543836236e-05
step: 390, loss: 2.805045915010851e-05
step: 400, loss: 3.746199217857793e-05
step: 410, loss: 3.116409425274469e-05
```

```

step: 420, loss: 3.072336039622314e-05
step: 430, loss: 3.233042298234068e-05
step: 440, loss: 1.5042569430079311e-05
step: 450, loss: 7.76993838371709e-05
step: 460, loss: 0.0001793199044186622
step: 470, loss: 1.5698193237767555e-05
step: 480, loss: 2.0064020645804703e-05
step: 490, loss: 2.566948205640074e-05
step: 500, loss: 3.0545790650648996e-05
step: 510, loss: 0.00024670333368703723
step: 520, loss: 4.377594086690806e-05
step: 530, loss: 5.602998498943634e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 14: dev_f1=0.9400749063670412, f1=0.9301025163094129,
best_f1=0.9301025163094129
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.1401299818535335e-05
step: 10, loss: 4.727256600745022e-05
step: 20, loss: 2.375548137933947e-05
step: 30, loss: 3.801038474193774e-05
step: 40, loss: 3.457986531429924e-05
step: 50, loss: 0.0009064654004760087
step: 60, loss: 1.9657898519653827e-05
step: 70, loss: 0.0004841288027819246
step: 80, loss: 2.7830734325107187e-05
step: 90, loss: 2.500245136616286e-05
step: 100, loss: 2.206068893428892e-05
step: 110, loss: 0.01655508577823639
step: 120, loss: 0.0018612104468047619
step: 130, loss: 1.9624321794253774e-05
step: 140, loss: 1.6521424186066724e-05
step: 150, loss: 0.0004660829436033964
step: 160, loss: 5.349096318241209e-05
step: 170, loss: 2.8869209927506745e-05
step: 180, loss: 1.987371979339514e-05
step: 190, loss: 7.016499148448929e-05
step: 200, loss: 3.755746365641244e-05
step: 210, loss: 4.473318767850287e-05
step: 220, loss: 8.443492697551847e-05
step: 230, loss: 0.00010278095578541979

```

```

step: 240, loss: 3.3082895242841914e-05
step: 250, loss: 3.839441342279315e-05
step: 260, loss: 2.6940035240841098e-05
step: 270, loss: 7.153535989345983e-05
step: 280, loss: 0.0008960620034486055
step: 290, loss: 0.0017141344724223018
step: 300, loss: 0.00023063315893523395
step: 310, loss: 4.366463326732628e-05
step: 320, loss: 3.5101023968309164e-05
step: 330, loss: 4.6147259126883e-05
step: 340, loss: 0.0002852007164619863
step: 350, loss: 1.817536394810304e-05
step: 360, loss: 3.0086845072219148e-05
step: 370, loss: 1.6946047253441066e-05
step: 380, loss: 1.6662963389535435e-05
step: 390, loss: 2.518162909836974e-05
step: 400, loss: 1.579869604029227e-05
step: 410, loss: 4.510895087150857e-05
step: 420, loss: 7.30337924323976e-05
step: 430, loss: 2.6657140551833436e-05
step: 440, loss: 0.00014601950533688068
step: 450, loss: 0.09102107584476471
step: 460, loss: 2.0089670215384103e-05
step: 470, loss: 4.602086846716702e-05
step: 480, loss: 2.6072104446939193e-05
step: 490, loss: 8.306180097861215e-05
step: 500, loss: 0.00020208772912155837
step: 510, loss: 2.2477819584310055e-05
step: 520, loss: 2.0682287868112326e-05
step: 530, loss: 2.8887559892609715e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9380028395646002, f1=0.9303857008466604,
best_f1=0.9301025163094129
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
```

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

5742it [00:14, 388.63it/s]

load\_f1 = 0.9380281690140844

real\_f1 = 0.9361502347417839

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

4398it [00:11, 372.82it/s]

### 3.1.4 Amazon-Google - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Amazon-Google \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize
```

```

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8307873606681824
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.0630994513630867
step: 20, loss: 0.3851255178451538
step: 30, loss: 0.3703988194465637
step: 40, loss: 0.49919310212135315
step: 50, loss: 0.29609623551368713
step: 60, loss: 0.3587312400341034
step: 70, loss: 0.22541989386081696
step: 80, loss: 0.2996882498264313
step: 90, loss: 0.3483390510082245
step: 100, loss: 0.11052104830741882
step: 110, loss: 0.24475383758544922
step: 120, loss: 0.22555069625377655
step: 130, loss: 0.24038800597190857
step: 140, loss: 0.21600918471813202
step: 150, loss: 0.2936384081840515
step: 160, loss: 0.32109421491622925
step: 170, loss: 0.1354384571313858
step: 180, loss: 0.1712019294500351
step: 190, loss: 0.22904515266418457
step: 200, loss: 0.15542279183864594
step: 210, loss: 0.5037827491760254
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6078098471986417, f1=0.5842293906810035,
best_f1=0.5842293906810035
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09191626310348511
step: 10, loss: 0.07561075687408447
step: 20, loss: 0.35239216685295105
step: 30, loss: 0.07891392707824707
step: 40, loss: 0.1573861986398697
step: 50, loss: 0.17647932469844818
step: 60, loss: 0.04436468333005905
step: 70, loss: 0.11648550629615784
step: 80, loss: 0.17511366307735443
step: 90, loss: 0.05346204340457916
step: 100, loss: 0.08164473623037338

```



```

step: 110, loss: 0.08834870159626007
step: 120, loss: 0.18582002818584442
step: 130, loss: 0.20945754647254944
step: 140, loss: 0.15019261837005615
step: 150, loss: 0.14478819072246552
step: 160, loss: 0.19996869564056396
step: 170, loss: 0.18953944742679596
step: 180, loss: 0.22646887600421906
step: 190, loss: 0.1957501918077469
step: 200, loss: 0.16075186431407928
step: 210, loss: 0.19049455225467682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6541353383458648, f1=0.6233269598470363,
best_f1=0.6233269598470363
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0957164317369461
step: 10, loss: 0.1697709560394287
step: 20, loss: 0.24471049010753632
step: 30, loss: 0.07394060492515564
step: 40, loss: 0.17013953626155853
step: 50, loss: 0.08443772047758102
step: 60, loss: 0.21843373775482178
step: 70, loss: 0.10375669598579407
step: 80, loss: 0.06542814522981644
step: 90, loss: 0.11018208414316177
step: 100, loss: 0.011987093836069107
step: 110, loss: 0.11476112902164459
step: 120, loss: 0.23294153809547424
step: 130, loss: 0.09409821778535843
step: 140, loss: 0.2236166000366211
step: 150, loss: 0.1930246353149414
step: 160, loss: 0.10513249039649963
step: 170, loss: 0.07328314334154129
step: 180, loss: 0.0673479363322258
step: 190, loss: 0.2721188962459564
step: 200, loss: 0.09554316848516464
step: 210, loss: 0.11442501097917557
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6642066420664207, f1=0.6275992438563327,
best_f1=0.6275992438563327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09965924918651581
step: 10, loss: 0.023929709568619728
step: 20, loss: 0.13295137882232666
step: 30, loss: 0.075421042740345
step: 40, loss: 0.10288075357675552
step: 50, loss: 0.03707844018936157
step: 60, loss: 0.06178131699562073
step: 70, loss: 0.3264651298522949
step: 80, loss: 0.1081276535987854
step: 90, loss: 0.06169070675969124
step: 100, loss: 0.054897040128707886
step: 110, loss: 0.1167331412434578
step: 120, loss: 0.024535011500120163
step: 130, loss: 0.030008958652615547
step: 140, loss: 0.13593274354934692
step: 150, loss: 0.1434197574853897
step: 160, loss: 0.020885499194264412
step: 170, loss: 0.03276173397898674
step: 180, loss: 0.2648155391216278
step: 190, loss: 0.06555303931236267
step: 200, loss: 0.04753749445080757
step: 210, loss: 0.010916316881775856
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.6817325800376649, f1=0.6457564575645757,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.054275210946798325
step: 10, loss: 0.075616754591465
step: 20, loss: 0.014307772740721703
step: 30, loss: 0.06687634438276291
step: 40, loss: 0.10434940457344055

```

```
step: 50, loss: 0.18553051352500916
step: 60, loss: 0.029030369594693184
step: 70, loss: 0.16618777811527252
step: 80, loss: 0.02443036623299122
step: 90, loss: 0.013150257989764214
step: 100, loss: 0.08386696875095367
step: 110, loss: 0.00403849920257926
step: 120, loss: 0.005846642889082432
step: 130, loss: 0.04204283282160759
step: 140, loss: 0.04282247647643089
step: 150, loss: 0.04976467415690422
step: 160, loss: 0.012315100058913231
step: 170, loss: 0.030877109616994858
step: 180, loss: 0.044392310082912445
step: 190, loss: 0.3191603124141693
step: 200, loss: 0.19018258154392242
step: 210, loss: 0.15481731295585632
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.6692913385826771, f1=0.6259842519685039,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0804155021905899
step: 10, loss: 0.03542422875761986
step: 20, loss: 0.054341595619916916
step: 30, loss: 0.0806884616613388
step: 40, loss: 0.031069105491042137
step: 50, loss: 0.02938215248286724
step: 60, loss: 0.1467818170785904
step: 70, loss: 0.0024262156803160906
step: 80, loss: 0.09633435308933258
step: 90, loss: 0.02736387774348259
step: 100, loss: 0.03517504781484604
step: 110, loss: 0.012393004260957241
step: 120, loss: 0.03526798263192177
step: 130, loss: 0.021519729867577553
step: 140, loss: 0.03312598913908005
step: 150, loss: 0.029904555529356003
step: 160, loss: 0.04627085104584694
step: 170, loss: 0.13198888301849365
step: 180, loss: 0.0018569286912679672
step: 190, loss: 0.021212086081504822
step: 200, loss: 0.015787115320563316
step: 210, loss: 0.0467098169028759
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6555555555555556, f1=0.6336283185840708,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014975606463849545
step: 10, loss: 0.007575994357466698
step: 20, loss: 0.01309833861887455
step: 30, loss: 0.005555913783609867
step: 40, loss: 0.056446272879838943
step: 50, loss: 0.04561896249651909
step: 60, loss: 0.19787512719631195
step: 70, loss: 0.0034368811175227165
step: 80, loss: 0.0394120030105114
step: 90, loss: 0.02429289184510708
step: 100, loss: 0.02648019790649414
step: 110, loss: 0.045737363398075104
step: 120, loss: 0.21325592696666718
step: 130, loss: 0.007049959152936935
step: 140, loss: 0.0016969001153483987
step: 150, loss: 0.020388588309288025
step: 160, loss: 0.005404782015830278
step: 170, loss: 0.020022325217723846
step: 180, loss: 0.0003377418324816972
step: 190, loss: 0.0017432058230042458
step: 200, loss: 0.010684813372790813
step: 210, loss: 0.04382391273975372
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.6460905349794239, f1=0.625, best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02927047200500965
step: 10, loss: 0.0024280049838125706
step: 20, loss: 0.017142662778496742
step: 30, loss: 0.05370137095451355
step: 40, loss: 0.017292503267526627
step: 50, loss: 0.09906252473592758
step: 60, loss: 0.09142877906560898
step: 70, loss: 0.0007799782324582338
step: 80, loss: 0.012668968178331852
step: 90, loss: 0.12103432416915894
step: 100, loss: 0.0273500457406044
step: 110, loss: 0.009294524788856506

```

```

step: 120, loss: 0.00521532166749239
step: 130, loss: 0.04301748052239418
step: 140, loss: 0.0038842938374727964
step: 150, loss: 0.004604909103363752
step: 160, loss: 0.05246601998806
step: 170, loss: 0.003550647059455514
step: 180, loss: 0.010098104365170002
step: 190, loss: 0.02646040730178356
step: 200, loss: 0.004159640055149794
step: 210, loss: 0.09721632301807404
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.650103519668737, f1=0.6467065868263474,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01864689029753208
step: 10, loss: 0.05749981477856636
step: 20, loss: 0.07128909975290298
step: 30, loss: 0.021357432007789612
step: 40, loss: 0.015079756267368793
step: 50, loss: 0.0007384063210338354
step: 60, loss: 0.012033249251544476
step: 70, loss: 0.05532991141080856
step: 80, loss: 0.0019768639467656612
step: 90, loss: 0.054367780685424805
step: 100, loss: 0.003008536295965314
step: 110, loss: 0.00033652488491497934
step: 120, loss: 0.0008114490192383528
step: 130, loss: 0.0009904816979542375
step: 140, loss: 0.023027591407299042
step: 150, loss: 0.001897274749353528
step: 160, loss: 0.01149099599570036
step: 170, loss: 0.04522942751646042
step: 180, loss: 0.04237031936645508
step: 190, loss: 0.030770964920520782
step: 200, loss: 0.02363712154328823
step: 210, loss: 0.03934966027736664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6455445544554456, f1=0.6463878326996197,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.0024374781642109156
step: 10, loss: 0.00044283008901402354
step: 20, loss: 0.004089963622391224
step: 30, loss: 0.0025405047927051783
step: 40, loss: 0.027064954861998558
step: 50, loss: 0.0009884153259918094
step: 60, loss: 0.00037096146843396127
step: 70, loss: 0.06744182854890823
step: 80, loss: 0.00010933431622106582
step: 90, loss: 0.0007623752462677658
step: 100, loss: 0.0010352643439546227
step: 110, loss: 0.005935326684266329
step: 120, loss: 0.015080937184393406
step: 130, loss: 0.00090134417405352
step: 140, loss: 0.0008606964838691056
step: 150, loss: 0.11431661993265152
step: 160, loss: 0.0385892428457737
step: 170, loss: 0.0037968566175550222
step: 180, loss: 0.03295384347438812
step: 190, loss: 0.011398895643651485
step: 200, loss: 0.030453095212578773
step: 210, loss: 0.012691996991634369
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6338582677165354, f1=0.631578947368421,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010086184367537498
step: 10, loss: 0.009778187610208988
step: 20, loss: 0.006255310028791428
step: 30, loss: 0.046579089015722275
step: 40, loss: 0.01464842353016138
step: 50, loss: 0.011351916939020157
step: 60, loss: 0.001983746187761426
step: 70, loss: 0.003781231353059411
step: 80, loss: 0.04237723723053932
step: 90, loss: 0.0006225704564712942
step: 100, loss: 0.0005129350465722382
step: 110, loss: 0.004648864269256592
step: 120, loss: 0.0007954750908538699
step: 130, loss: 0.061373259872198105
step: 140, loss: 0.0031431401148438454
step: 150, loss: 0.09324024617671967
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.0027445233426988125

```

```

step: 170, loss: 0.1302938461303711
step: 180, loss: 0.015684030950069427
step: 190, loss: 0.0004927123663946986
step: 200, loss: 0.00023273422266356647
step: 210, loss: 0.04858899489045143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6519114688128774, f1=0.6461538461538461,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011207886273041368
step: 10, loss: 0.006700923666357994
step: 20, loss: 0.0030647115781903267
step: 30, loss: 0.012120133265852928
step: 40, loss: 0.0015648460248485208
step: 50, loss: 0.11963552981615067
step: 60, loss: 0.00028382588061504066
step: 70, loss: 0.0008990815840661526
step: 80, loss: 0.000675519579090178
step: 90, loss: 0.001214375370182097
step: 100, loss: 0.01098909042775631
step: 110, loss: 0.002398877404630184
step: 120, loss: 0.05002665892243385
step: 130, loss: 0.003719300962984562
step: 140, loss: 0.0012796939117833972
step: 150, loss: 0.0009800954721868038
step: 160, loss: 0.0011150211794301867
step: 170, loss: 0.00041124477866105735
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 180, loss: 0.0010488391853868961
step: 190, loss: 0.045627571642398834
step: 200, loss: 0.082797110080719
step: 210, loss: 0.0029997434467077255
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6394849785407726, f1=0.6317991631799162,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013101960939820856
step: 10, loss: 0.0035843071527779102
step: 20, loss: 0.0002613621181808412
step: 30, loss: 0.012801934964954853

```

```

step: 40, loss: 0.00045274445437826216
step: 50, loss: 0.0001355817075818777
step: 60, loss: 0.010451885871589184
step: 70, loss: 0.0005922044510953128
step: 80, loss: 0.01772548444569111
step: 90, loss: 0.0361938513815403
step: 100, loss: 0.030832655727863312
step: 110, loss: 0.0002631679526530206
step: 120, loss: 0.0023235161788761616
step: 130, loss: 0.007428164593875408
step: 140, loss: 0.0009140464244410396
step: 150, loss: 0.000469276710646227
step: 160, loss: 0.0013090670108795166
step: 170, loss: 0.0009539643069729209
step: 180, loss: 0.03803089261054993
step: 190, loss: 0.0004470533167477697
step: 200, loss: 0.00203684251755476
step: 210, loss: 0.011375539004802704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6452905811623246, f1=0.6490566037735849,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00996454805135727
step: 10, loss: 0.0004028955299872905
step: 20, loss: 0.0003135353035759181
step: 30, loss: 0.00042141310404986143
step: 40, loss: 0.007980464957654476
step: 50, loss: 0.002813121769577265
step: 60, loss: 0.00021014467347413301
step: 70, loss: 0.00153937132563442
step: 80, loss: 0.07279355078935623
step: 90, loss: 0.00043009952059946954
step: 100, loss: 0.001300811767578125
step: 110, loss: 0.007398050744086504
step: 120, loss: 0.00023375486489385366
step: 130, loss: 0.00013511627912521362
step: 140, loss: 0.0002678293676581234
step: 150, loss: 0.014691133983433247
step: 160, loss: 0.004831704311072826
step: 170, loss: 0.010741294361650944
step: 180, loss: 0.011832760646939278
step: 190, loss: 0.00016751963994465768
step: 200, loss: 0.00021859732805751264
step: 210, loss: 0.0005154259270057082

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6460176991150443, f1=0.6048565121412804,
best_f1=0.6457564575645757
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01720587909221649
step: 10, loss: 0.0034833760000765324
step: 20, loss: 0.00016822559700813144
step: 30, loss: 0.00020289378880988806
step: 40, loss: 0.00018068474309984595
step: 50, loss: 0.00015097016876097769
step: 60, loss: 0.0011731042759492993
step: 70, loss: 0.0005386452539823949
step: 80, loss: 0.00046182499499991536
step: 90, loss: 0.034011777490377426
step: 100, loss: 0.0006609167903661728
step: 110, loss: 0.00010301404108759016
step: 120, loss: 0.0006884722970426083
step: 130, loss: 0.0002978515694849193
step: 140, loss: 0.052714753895998
step: 150, loss: 0.00024236684839706868
step: 160, loss: 0.08836045116186142
step: 170, loss: 0.011628824286162853
step: 180, loss: 0.0013650491600856185
step: 190, loss: 0.0021859160624444485
step: 200, loss: 0.04087357223033905
step: 210, loss: 0.00041502705425955355
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6478260869565218, f1=0.6209850107066381,
best_f1=0.6457564575645757
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'") /usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

2293it [00:03, 649.85it/s]

load\_f1 = 0.6574803149606299

real\_f1 = 0.6533066132264529

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

4398it [00:10, 407.62it/s]

### 3.1.5 Walmart-Amazon - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Walmart-Amazon \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize
```

```

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Walmart-Amazon \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
 /usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
 /usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8466381430625916
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16198042035102844
step: 20, loss: 0.15038460493087769
step: 30, loss: 0.5061901807785034
step: 40, loss: 0.2590300738811493
step: 50, loss: 0.3092843294143677
step: 60, loss: 0.35709258913993835
step: 70, loss: 0.17731401324272156
step: 80, loss: 0.5233616828918457
step: 90, loss: 0.23955149948596954
step: 100, loss: 0.22011423110961914
step: 110, loss: 0.2343457043170929
step: 120, loss: 0.4108676016330719
step: 130, loss: 0.3377395570278168
step: 140, loss: 0.3231295049190521
step: 150, loss: 0.24992682039737701
step: 160, loss: 0.20855718851089478
step: 170, loss: 0.38893216848373413
step: 180, loss: 0.27587631344795227
step: 190, loss: 0.1300353854894638
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6153846153846154, f1=0.5590062111801243,
best_f1=0.5590062111801243
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2282181680202484
step: 10, loss: 0.052401699125766754
step: 20, loss: 0.07370320707559586
step: 30, loss: 0.17142271995544434
step: 40, loss: 0.38889050483703613
step: 50, loss: 0.31504786014556885
step: 60, loss: 0.24416302144527435
step: 70, loss: 0.1766684353351593
step: 80, loss: 0.13467822968959808
step: 90, loss: 0.10864567756652832
step: 100, loss: 0.26769086718559265
step: 110, loss: 0.19201791286468506
step: 120, loss: 0.10482759028673172

```

```

step: 130, loss: 0.053974322974681854
step: 140, loss: 0.1743118166923523
step: 150, loss: 0.062003929167985916
step: 160, loss: 0.09145323187112808
step: 170, loss: 0.2245255708694458
step: 180, loss: 0.08153665065765381
step: 190, loss: 0.06712137907743454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7272727272727272, f1=0.7038123167155425,
best_f1=0.7038123167155425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11072459071874619
step: 10, loss: 0.21563120186328888
step: 20, loss: 0.12238971143960953
step: 30, loss: 0.035808831453323364
step: 40, loss: 0.05428750813007355
step: 50, loss: 0.08054625988006592
step: 60, loss: 0.08525906503200531
step: 70, loss: 0.13535206019878387
step: 80, loss: 0.13609525561332703
step: 90, loss: 0.06205115094780922
step: 100, loss: 0.125777930021286
step: 110, loss: 0.0975990742444992
step: 120, loss: 0.03275677189230919
step: 130, loss: 0.16989606618881226
step: 140, loss: 0.039733897894620895
step: 150, loss: 0.03280627727508545
step: 160, loss: 0.15538626909255981
step: 170, loss: 0.13832305371761322
step: 180, loss: 0.014614222571253777
step: 190, loss: 0.14167138934135437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7720207253886011, f1=0.774869109947644,
best_f1=0.774869109947644

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017899462953209877
step: 10, loss: 0.03639570623636246
step: 20, loss: 0.017753448337316513
step: 30, loss: 0.08143807202577591
step: 40, loss: 0.0030636871233582497
step: 50, loss: 0.024855291470885277
step: 60, loss: 0.12940984964370728
step: 70, loss: 0.002758026123046875
step: 80, loss: 0.02243603579699993
step: 90, loss: 0.026033444330096245
step: 100, loss: 0.009313282556831837
step: 110, loss: 0.0036085378378629684
step: 120, loss: 0.17848435044288635
step: 130, loss: 0.2395259439945221
step: 140, loss: 0.05040832981467247
step: 150, loss: 0.10806379467248917
step: 160, loss: 0.07160468399524689
step: 170, loss: 0.005075359251350164
step: 180, loss: 0.08068128675222397
step: 190, loss: 0.07105065137147903
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.7704918032786885, f1=0.760806916426513,
best_f1=0.774869109947644
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0071211764588952065
step: 10, loss: 0.015716660767793655
step: 20, loss: 0.06745714694261551
step: 30, loss: 0.007976355962455273
step: 40, loss: 0.10423830151557922
step: 50, loss: 0.08925191313028336
step: 60, loss: 0.01432829536497593
step: 70, loss: 0.0011457083746790886
step: 80, loss: 0.008622676134109497
step: 90, loss: 0.006841168273240328
step: 100, loss: 0.057942621409893036
step: 110, loss: 0.00336740305647254
step: 120, loss: 0.0009246262488886714
step: 130, loss: 0.12116480618715286
step: 140, loss: 0.005224199965596199
step: 150, loss: 0.016297876834869385
step: 160, loss: 0.002331108320504427

```

```

step: 170, loss: 0.031987521797418594
step: 180, loss: 0.03809470683336258
step: 190, loss: 0.14751611649990082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7777777777777779, f1=0.7520435967302452,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11266357451677322
step: 10, loss: 0.023592885583639145
step: 20, loss: 0.00336557999253273
step: 30, loss: 0.0012132682604715228
step: 40, loss: 0.0031332599464803934
step: 50, loss: 0.0020394036546349525
step: 60, loss: 0.0006644679815508425
step: 70, loss: 0.005268275737762451
step: 80, loss: 0.10062223672866821
step: 90, loss: 0.02887612394988537
step: 100, loss: 0.0022906060330569744
step: 110, loss: 0.015931066125631332
step: 120, loss: 0.1109333336353302
step: 130, loss: 0.03718892112374306
step: 140, loss: 0.0029884600080549717
step: 150, loss: 0.08395872265100479
step: 160, loss: 0.005985778756439686
step: 170, loss: 0.03646137937903404
step: 180, loss: 0.022093558683991432
step: 190, loss: 0.013737078756093979
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.773067331670823, f1=0.7409326424870466,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028519509360194206
step: 10, loss: 0.013694736175239086
step: 20, loss: 0.0018076077103614807
step: 30, loss: 0.004742997232824564
step: 40, loss: 0.006010239012539387

```

```

step: 50, loss: 0.03596414253115654
step: 60, loss: 0.004975615534931421
step: 70, loss: 0.0005527967005036771
step: 80, loss: 0.017191261053085327
step: 90, loss: 0.0011313142022117972
step: 100, loss: 0.029639268293976784
step: 110, loss: 0.018704283982515335
step: 120, loss: 0.006105133332312107
step: 130, loss: 0.001838410273194313
step: 140, loss: 0.0005423919064924121
step: 150, loss: 0.0031245555728673935
step: 160, loss: 0.0011167583288624883
step: 170, loss: 0.003952597267925739
step: 180, loss: 0.001853077206760645
step: 190, loss: 0.04962647333741188
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.762836185819071, f1=0.7218045112781954,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005296029266901314
step: 10, loss: 0.00293682306073606
step: 20, loss: 0.0008171623921953142
step: 30, loss: 0.005922022741287947
step: 40, loss: 0.0012284655822440982
step: 50, loss: 0.0026244367472827435
step: 60, loss: 0.0005522173596546054
step: 70, loss: 0.00045980530558153987
step: 80, loss: 0.012221257202327251
step: 90, loss: 0.004274051170796156
step: 100, loss: 0.02490011602640152
step: 110, loss: 0.0007624649442732334
step: 120, loss: 0.0007367050857283175
step: 130, loss: 0.000950864574406296
step: 140, loss: 0.010325465351343155
step: 150, loss: 0.04058994725346565
step: 160, loss: 0.0033152096439152956
step: 170, loss: 0.00038178058457560837
step: 180, loss: 0.0022863815538585186
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 190, loss: 0.0013753945240750909
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7334963325183375, f1=0.7376237623762376,

```



```

best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006842180737294257
step: 10, loss: 0.002065549837425351
step: 20, loss: 0.020712237805128098
step: 30, loss: 0.017756300047039986
step: 40, loss: 0.14118966460227966
step: 50, loss: 0.00040401588194072247
step: 60, loss: 0.001445602043531835
step: 70, loss: 0.07081782817840576
step: 80, loss: 0.001112213358283043
step: 90, loss: 0.020257627591490746
step: 100, loss: 0.011983130127191544
step: 110, loss: 0.04829394072294235
step: 120, loss: 0.019605303183197975
step: 130, loss: 0.008113578893244267
step: 140, loss: 0.009469144977629185
step: 150, loss: 0.056863874197006226
step: 160, loss: 0.09632240235805511
step: 170, loss: 0.00364032038487494
step: 180, loss: 0.002606655703857541
step: 190, loss: 0.0005902368575334549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7512437810945273, f1=0.7277353689567431,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006326761795207858
step: 10, loss: 0.00095456896815449
step: 20, loss: 0.0010551951127126813
step: 30, loss: 0.0004969558794982731
step: 40, loss: 0.0004655302909668535
step: 50, loss: 0.00032587823807261884
step: 60, loss: 0.009659016504883766
step: 70, loss: 0.00250012194737792
step: 80, loss: 0.0002235587453469634
step: 90, loss: 0.00504083838313818
step: 100, loss: 0.0005637871799990535
step: 110, loss: 0.0453752800822258
step: 120, loss: 0.010793675668537617
step: 130, loss: 0.001002511358819902
step: 140, loss: 0.11579977720975876
step: 150, loss: 0.0002318637416465208

```

```

step: 160, loss: 0.005837453529238701
step: 170, loss: 0.0011573823867365718
step: 180, loss: 0.00041746938950382173
step: 190, loss: 0.0007950537838041782
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7611548556430445, f1=0.745308310991957,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006010690703988075
step: 10, loss: 0.0007720050634816289
step: 20, loss: 0.0006673837197013199
step: 30, loss: 0.0019287597388029099
step: 40, loss: 0.003726917551830411
step: 50, loss: 0.00039692208520136774
step: 60, loss: 0.009127249009907246
step: 70, loss: 0.0012041863519698381
step: 80, loss: 0.0002765073149930686
step: 90, loss: 0.0027236947789788246
step: 100, loss: 0.0006547273369506001
step: 110, loss: 0.0004864154616370797
step: 120, loss: 0.0018579141469672322
step: 130, loss: 0.00034845576738007367
step: 140, loss: 0.00017807558469939977
step: 150, loss: 0.0004620729887392372
step: 160, loss: 0.0019136620685458183
step: 170, loss: 0.000701634562574327
step: 180, loss: 0.0012869744095951319
step: 190, loss: 0.0007310583023354411
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.767123287671233, f1=0.7541899441340784,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00845650676637888
step: 10, loss: 0.0018901810981333256
step: 20, loss: 0.008807909674942493
step: 30, loss: 0.15058952569961548
step: 40, loss: 0.0004477968323044479
step: 50, loss: 0.0007990024168975651
step: 60, loss: 0.0006923836772330105
step: 70, loss: 0.000471185747301206

```

```

step: 80, loss: 0.0003725796123035252
step: 90, loss: 0.0013528052950277925
step: 100, loss: 0.0004283181915525347
step: 110, loss: 0.008573794737458229
step: 120, loss: 0.00030117519781924784
step: 130, loss: 0.00021505057520698756
step: 140, loss: 0.006042852532118559
step: 150, loss: 0.0005213020485825837
step: 160, loss: 0.0004014832084067166
step: 170, loss: 0.04487176984548569
step: 180, loss: 0.00024253579613287002
step: 190, loss: 0.0023409537971019745
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.766304347826087, f1=0.7450980392156863,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000601018313318491
step: 10, loss: 0.000858549028635025
step: 20, loss: 0.0008397457422688603
step: 30, loss: 0.0007396453293040395
step: 40, loss: 0.00010125287371920422
step: 50, loss: 0.001038455287925899
step: 60, loss: 0.00016883376520127058
step: 70, loss: 0.0002546449250075966
step: 80, loss: 0.0002222758048446849
step: 90, loss: 0.000266836752416566
step: 100, loss: 0.00030416686786338687
step: 110, loss: 0.0366412028670311
step: 120, loss: 0.009014037437736988
step: 130, loss: 0.0007723078597337008
step: 140, loss: 0.00024792790645733476
step: 150, loss: 0.0003283985424786806
step: 160, loss: 0.0002882717817556113
step: 170, loss: 7.24776546121575e-05
step: 180, loss: 0.007659607101231813
step: 190, loss: 0.00047514107427559793
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7608695652173914, f1=0.7430167597765364,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.0005744681693613529
step: 10, loss: 0.0009251185692846775
step: 20, loss: 0.00023441831581294537
step: 30, loss: 0.08074740320444107
step: 40, loss: 0.00028009183006361127
step: 50, loss: 0.01491521019488573
step: 60, loss: 0.0001925575634231791
step: 70, loss: 0.00403823284432292
step: 80, loss: 0.0001695876708254218
step: 90, loss: 0.0012774475617334247
step: 100, loss: 0.000514127139467746
step: 110, loss: 0.002483016811311245
step: 120, loss: 0.000198079040274024
step: 130, loss: 0.00013596749340649694
step: 140, loss: 0.000131781940581277
step: 150, loss: 0.00021801138063892722
step: 160, loss: 0.0026586377061903477
step: 170, loss: 0.0004994146293029189
step: 180, loss: 0.00037113201688043773
step: 190, loss: 8.971740317065269e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7647058823529412, f1=0.7439353099730459,
best_f1=0.7520435967302452
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000987380393780768
step: 10, loss: 0.0002867550647351891
step: 20, loss: 0.00045468437019735575
step: 30, loss: 0.00022770381474401802
step: 40, loss: 0.00019271338533144444
step: 50, loss: 0.011092070490121841
step: 60, loss: 0.00040897284634411335
step: 70, loss: 0.00018343504052609205
step: 80, loss: 0.00011013530456693843
step: 90, loss: 0.0019885036163032055
step: 100, loss: 0.0006656938348896801
step: 110, loss: 0.00016280171985272318
step: 120, loss: 0.00035129504976794124
step: 130, loss: 0.0003793981741182506
step: 140, loss: 0.00023629060888197273
step: 150, loss: 0.0006048720679245889
step: 160, loss: 0.0007213260978460312
step: 170, loss: 0.0008443117840215564
step: 180, loss: 0.004113762639462948
step: 190, loss: 0.004421195946633816

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7647058823529412, f1=0.7540983606557378,
best_f1=0.7520435967302452
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:06, 335.80it/s]
load_f1 = 0.6324786324786325
real_f1 = 0.6010638297872342
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 398.90it/s]

```

## 3.2 DK TEXTUAL

### 3.2.1 Abt-Buy - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Textual/Abt-Buy \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Textual/Abt-Buy \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8502892255783081
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22505833208560944
step: 20, loss: 0.15758077800273895
step: 30, loss: 0.2256447821855545
step: 40, loss: 0.3146975040435791
step: 50, loss: 0.3795931935310364
step: 60, loss: 0.44684910774230957
step: 70, loss: 0.30678263306617737
step: 80, loss: 0.2482096552848816
step: 90, loss: 0.40745311975479126
step: 100, loss: 0.23843809962272644
step: 110, loss: 0.185938760638237
step: 120, loss: 0.5241210460662842
step: 130, loss: 0.4142914414405823
step: 140, loss: 0.4568012058734894
step: 150, loss: 0.08904366195201874
step: 160, loss: 0.23517850041389465
step: 170, loss: 0.123810775578022
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 1: dev\_f1=0.6774193548387097, f1=0.6837209302325582,

best\_f1=0.6837209302325582

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.2803961932659149

step: 10, loss: 0.05723877251148224

step: 20, loss: 0.2450023889541626

step: 30, loss: 0.17329108715057373

step: 40, loss: 0.14459159970283508

step: 50, loss: 0.17130500078201294

step: 60, loss: 0.11341013759374619

step: 70, loss: 0.1834072470664978

step: 80, loss: 0.16590969264507294

step: 90, loss: 0.15348809957504272

step: 100, loss: 0.11278008669614792

step: 110, loss: 0.18036988377571106

step: 120, loss: 0.03457151725888252

step: 130, loss: 0.05964047461748123

step: 140, loss: 0.17605552077293396

step: 150, loss: 0.07248065620660782

step: 160, loss: 0.15929588675498962

step: 170, loss: 0.03619295731186867

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 2: dev\_f1=0.7339901477832513, f1=0.7488151658767772,

best\_f1=0.7488151658767772

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.07138276845216751

step: 10, loss: 0.03486404940485954

step: 20, loss: 0.16607099771499634

step: 30, loss: 0.06862220168113708

step: 40, loss: 0.008144695311784744

step: 50, loss: 0.10366176068782806

step: 60, loss: 0.09669926762580872

step: 70, loss: 0.04727323353290558

step: 80, loss: 0.1722577065229416

step: 90, loss: 0.1194271445274353



```

step: 100, loss: 0.01749366708099842
step: 110, loss: 0.04363778606057167
step: 120, loss: 0.00399720948189497
step: 130, loss: 0.09729502350091934
step: 140, loss: 0.0025787141639739275
step: 150, loss: 0.0410454235970974
step: 160, loss: 0.04584748297929764
step: 170, loss: 0.08973380923271179
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7714987714987714, f1=0.7772511848341233,
best_f1=0.7772511848341233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06407091021537781
step: 10, loss: 0.1764160692691803
step: 20, loss: 0.01628628931939602
step: 30, loss: 0.024892833083868027
step: 40, loss: 0.009925952181220055
step: 50, loss: 0.011903936043381691
step: 60, loss: 0.07648568600416183
step: 70, loss: 0.01803439110517502
step: 80, loss: 0.09451266378164291
step: 90, loss: 0.031169403344392776
step: 100, loss: 0.0027702294755727053
step: 110, loss: 0.02287449687719345
step: 120, loss: 0.1094602420926094
step: 130, loss: 0.0750567689538002
step: 140, loss: 0.026052257046103477
step: 150, loss: 0.14937472343444824
step: 160, loss: 0.07184053957462311
step: 170, loss: 0.010866356082260609
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7911547911547911, f1=0.7889908256880733,
best_f1=0.7889908256880733
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04067192226648331
step: 10, loss: 0.012351181358098984
step: 20, loss: 0.02087380923330784
step: 30, loss: 0.03537127375602722
step: 40, loss: 0.020097002387046814
step: 50, loss: 0.02040434628725052
step: 60, loss: 0.0024335887283086777
step: 70, loss: 0.013579800724983215
step: 80, loss: 0.015256659127771854
step: 90, loss: 0.12484823912382126
step: 100, loss: 0.015461592003703117
step: 110, loss: 0.22393012046813965
step: 120, loss: 0.007261296734213829
step: 130, loss: 0.0038779897149652243
step: 140, loss: 0.019640779122710228
step: 150, loss: 0.07735355943441391
step: 160, loss: 0.010666340589523315
step: 170, loss: 0.059614766389131546
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.8050632911392405, f1=0.7931873479318734,
best_f1=0.7931873479318734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028650788590312004
step: 10, loss: 0.058620184659957886
step: 20, loss: 0.015808068215847015
step: 30, loss: 0.01248860266059637
step: 40, loss: 0.0017491834005340934
step: 50, loss: 0.012634824961423874
step: 60, loss: 0.027072139084339142
step: 70, loss: 0.0326317735016346
step: 80, loss: 0.036801777780056
step: 90, loss: 0.07729720324277878
step: 100, loss: 0.003616112284362316
step: 110, loss: 0.02143007330596447
step: 120, loss: 0.025617225095629692
step: 130, loss: 0.1758299320936203
step: 140, loss: 0.16122674942016602
step: 150, loss: 0.10448675602674484

```

```

step: 160, loss: 0.010553943924605846
step: 170, loss: 0.0035754593554884195
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7895981087470448, f1=0.7803738317757009,
best_f1=0.7931873479318734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03374500572681427
step: 10, loss: 0.0029632344376295805
step: 20, loss: 0.04934677109122276
step: 30, loss: 0.027018534019589424
step: 40, loss: 0.014180084690451622
step: 50, loss: 0.0010515146423131227
step: 60, loss: 0.15065082907676697
step: 70, loss: 0.0071213748306035995
step: 80, loss: 0.0008258839370682836
step: 90, loss: 0.0040435888804495335
step: 100, loss: 0.09740456938743591
step: 110, loss: 0.0009070377564057708
step: 120, loss: 0.06218405440449715
step: 130, loss: 0.11391060054302216
step: 140, loss: 0.01379632018506527
step: 150, loss: 0.02549072913825512
step: 160, loss: 0.009993068873882294
step: 170, loss: 0.28026461601257324
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.8058252427184465, f1=0.7848699763593381,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02567896991968155
step: 10, loss: 0.00351475621573627
step: 20, loss: 0.027891865000128746
step: 30, loss: 0.004800410475581884
step: 40, loss: 0.0038846186362206936
step: 50, loss: 0.009274665266275406
step: 60, loss: 0.010734452866017818
step: 70, loss: 0.009117810055613518

```

```

step: 80, loss: 0.0007651274208910763
step: 90, loss: 0.06861262023448944
step: 100, loss: 0.013286639004945755
step: 110, loss: 0.012675434350967407
step: 120, loss: 0.005323120858520269
step: 130, loss: 0.0009794129291549325
step: 140, loss: 0.001984661677852273
step: 150, loss: 0.03249288722872734
step: 160, loss: 0.12714362144470215
step: 170, loss: 0.0305876936763525
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7780548628428928, f1=0.7894736842105263,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0031327565666288137
step: 10, loss: 0.16689279675483704
step: 20, loss: 0.002816753927618265
step: 30, loss: 0.01378579530864954
step: 40, loss: 0.01511465199291706
step: 50, loss: 0.01107370387762785
step: 60, loss: 0.001953297993168235
step: 70, loss: 0.03263087570667267
step: 80, loss: 0.04878681153059006
step: 90, loss: 0.0020405035465955734
step: 100, loss: 0.02882865071296692
step: 110, loss: 0.00029319984605535865
step: 120, loss: 0.014242672361433506
step: 130, loss: 0.0008056049118749797
step: 140, loss: 0.0009014755487442017
step: 150, loss: 0.003319605952128768
step: 160, loss: 0.005091393366456032
step: 170, loss: 0.006843088194727898
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7897435897435897, f1=0.7796610169491525,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008812425658106804
step: 10, loss: 0.01952725648880005
step: 20, loss: 0.0011267693480476737
step: 30, loss: 0.0705680102109909

```

```

step: 40, loss: 0.0016571249580010772
step: 50, loss: 0.006003225687891245
step: 60, loss: 0.0009227333357557654
step: 70, loss: 0.04445138946175575
step: 80, loss: 0.03815402090549469
step: 90, loss: 0.09859524667263031
step: 100, loss: 0.006460196804255247
step: 110, loss: 0.0004112109018024057
step: 120, loss: 0.016188016161322594
step: 130, loss: 0.009230561554431915
step: 140, loss: 0.03562181442975998
step: 150, loss: 0.0007058957708068192
step: 160, loss: 0.0011280336184427142
step: 170, loss: 0.0016661619301885366
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7980295566502462, f1=0.7848699763593381,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006656942423433065
step: 10, loss: 0.0170600488781929
step: 20, loss: 0.013401924632489681
step: 30, loss: 0.0011458336375653744
step: 40, loss: 0.0027093542739748955
step: 50, loss: 0.00037724722642451525
step: 60, loss: 0.0005428250879049301
step: 70, loss: 0.0021525979973375797
step: 80, loss: 0.0006335952784866095
step: 90, loss: 0.0004149048472754657
step: 100, loss: 0.0006416841060854495
step: 110, loss: 0.001980353379622102
step: 120, loss: 0.0017080975230783224
step: 130, loss: 0.00089462153846398
step: 140, loss: 0.031944841146469116
step: 150, loss: 0.0006915358826518059
step: 160, loss: 0.0002581427397672087
step: 170, loss: 0.018227681517601013
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7930174563591021, f1=0.7877358490566038,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.0011102259159088135
step: 10, loss: 0.005099388770759106
step: 20, loss: 0.02337256819009781
step: 30, loss: 0.0007177255465649068
step: 40, loss: 0.0005186707712709904
step: 50, loss: 0.0018369107274338603
step: 60, loss: 0.0006321757100522518
step: 70, loss: 0.015038929879665375
step: 80, loss: 0.019087238237261772
step: 90, loss: 0.00039231759728863835
step: 100, loss: 0.0069253211840987206
step: 110, loss: 0.06721431016921997
step: 120, loss: 0.027861809358000755
step: 130, loss: 0.0018308070721104741
step: 140, loss: 0.03117303177714348
step: 150, loss: 0.0009448732598684728
step: 160, loss: 0.014728658832609653
step: 170, loss: 0.15936008095741272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7927461139896372, f1=0.794188861985472,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0053618368692696095
step: 10, loss: 0.019443826749920845
step: 20, loss: 0.0006349190953187644
step: 30, loss: 0.0014786302344873548
step: 40, loss: 0.00036341059603728354
step: 50, loss: 0.000976642593741417
step: 60, loss: 0.0016826885985210538
step: 70, loss: 0.0006652632146142423
step: 80, loss: 0.0002623506879899651
step: 90, loss: 0.00311425793915987
step: 100, loss: 0.00018013108638115227
step: 110, loss: 0.001940157380886376
step: 120, loss: 0.002784343436360359
step: 130, loss: 0.000328750436892733
step: 140, loss: 0.0005827353452332318
step: 150, loss: 0.030669644474983215
step: 160, loss: 0.0009842708241194487
step: 170, loss: 0.0008556881221011281
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7938931297709925, f1=0.7933491686460807,

```

```

best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0048733120784163475
step: 10, loss: 0.00037740293191745877
step: 20, loss: 0.004789819475263357
step: 30, loss: 0.0029826427344232798
step: 40, loss: 0.00041311673703603446
step: 50, loss: 0.0042958445847034454
step: 60, loss: 0.00044399723992682993
step: 70, loss: 0.0017139388946816325
step: 80, loss: 0.0019885601941496134
step: 90, loss: 0.0002592475793790072
step: 100, loss: 0.00021191693667788059
step: 110, loss: 0.0038033882156014442
step: 120, loss: 0.0038836633320897818
step: 130, loss: 0.00020814903837163
step: 140, loss: 0.00045118556590750813
step: 150, loss: 0.0022683637216687202
step: 160, loss: 0.03741973638534546
step: 170, loss: 0.0005212606047280133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7990074441687345, f1=0.7887323943661971,
best_f1=0.7848699763593381
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014033474726602435
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.004537749569863081
step: 20, loss: 0.00847350899130106
step: 30, loss: 0.01706628128886223
step: 40, loss: 0.0002878323430195451
step: 50, loss: 0.0002707981620915234
step: 60, loss: 0.00011042052938137203
step: 70, loss: 0.0032481951639056206
step: 80, loss: 0.00015459544374607503
step: 90, loss: 0.00011600498692132533
step: 100, loss: 0.0002944708103314042
step: 110, loss: 0.00014514909707941115
step: 120, loss: 0.0033299496863037348
step: 130, loss: 0.00035957733052782714
step: 140, loss: 0.00020477383804973215
step: 150, loss: 0.0021418496035039425
step: 160, loss: 0.00031761921127326787

```

```

step: 170, loss: 0.0006297711515799165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.8009950248756218, f1=0.7905882352941176,
best_f1=0.7848699763593381
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:04, 437.15it/s]
load_f1 = 0.4833948339483395
real_f1 = 0.4440894568690096
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 402.43it/s]

```



### 3.3 DK DIRTY

#### 3.3.1 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Dirty/DBLP-ACM \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Dirty/DBLP-ACM \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7934724688529968
step: 10, loss: 0.4671390652656555
step: 20, loss: 0.5591360330581665
step: 30, loss: 0.34424686431884766
step: 40, loss: 0.12439101934432983
step: 50, loss: 0.1695077121257782
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.05248429998755455
step: 70, loss: 0.15827777981758118
step: 80, loss: 0.20654846727848053
step: 90, loss: 0.038042958825826645
step: 100, loss: 0.09139401465654373
step: 110, loss: 0.11114636808633804
step: 120, loss: 0.02170238457620144
step: 130, loss: 0.008890178054571152
step: 140, loss: 0.0106574771925807
step: 150, loss: 0.16174453496932983
step: 160, loss: 0.2368292659521103
step: 170, loss: 0.014330344274640083
step: 180, loss: 0.011499704793095589
step: 190, loss: 0.007525652647018433
step: 200, loss: 0.008290410041809082
step: 210, loss: 0.0071871778927743435

```

```

step: 220, loss: 0.04395383223891258
step: 230, loss: 0.024562276899814606
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9609810479375697, f1=0.9574944071588367,
best_f1=0.9574944071588367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014457745477557182
step: 10, loss: 0.01340015884488821
step: 20, loss: 0.008438199758529663
step: 30, loss: 0.004004491958767176
step: 40, loss: 0.011507458984851837
step: 50, loss: 0.0050668371841311455
step: 60, loss: 0.01729806698858738
step: 70, loss: 0.2074832171201706
step: 80, loss: 0.0077224913984537125
step: 90, loss: 0.007554491516202688
step: 100, loss: 0.09749439358711243
step: 110, loss: 0.13857638835906982
step: 120, loss: 0.006760159507393837
step: 130, loss: 0.01442879717797041
step: 140, loss: 0.13905394077301025
step: 150, loss: 0.013306383974850178
step: 160, loss: 0.006403014063835144
step: 170, loss: 0.012590261176228523
step: 180, loss: 0.006647519301623106
step: 190, loss: 0.08723543584346771
step: 200, loss: 0.0047095962800085545
step: 210, loss: 0.06264925003051758
step: 220, loss: 0.0015497299609705806
step: 230, loss: 0.1144329234957695
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.972972972972973, f1=0.9628796400449944,
best_f1=0.9628796400449944
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022753795608878136
step: 10, loss: 0.013012668117880821
step: 20, loss: 0.004878098610788584
step: 30, loss: 0.021114850416779518
step: 40, loss: 0.01851622387766838
step: 50, loss: 0.009490749798715115
step: 60, loss: 0.09596389532089233
step: 70, loss: 0.0027712832670658827
step: 80, loss: 0.06130403280258179
step: 90, loss: 0.1247156634926796
step: 100, loss: 0.018430881202220917
step: 110, loss: 0.00833277590572834
step: 120, loss: 0.009878380224108696
step: 130, loss: 0.00057123729493469
step: 140, loss: 0.001370170502923429
step: 150, loss: 0.0010815071873366833
step: 160, loss: 0.0017885133856907487
step: 170, loss: 0.016113433986902237
step: 180, loss: 0.0056467484682798386
step: 190, loss: 0.00377086759544909
step: 200, loss: 0.007463404908776283
step: 210, loss: 0.02296920120716095
step: 220, loss: 0.05723315849900246
step: 230, loss: 0.012174061499536037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9743016759776536, f1=0.9663677130044843,
best_f1=0.9663677130044843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013085214421153069
step: 10, loss: 0.005124371033161879
step: 20, loss: 0.0013496269239112735
step: 30, loss: 0.003295990638434887
step: 40, loss: 0.009457048960030079
step: 50, loss: 0.0034485834185034037
step: 60, loss: 0.00048767554108053446
step: 70, loss: 0.010218084789812565
step: 80, loss: 0.031247539445757866
step: 90, loss: 0.022051895037293434

```

```

step: 100, loss: 0.0010883131762966514
step: 110, loss: 0.005167654249817133
step: 120, loss: 0.008584028109908104
step: 130, loss: 0.045201100409030914
step: 140, loss: 0.0007271667709574103
step: 150, loss: 0.004632999654859304
step: 160, loss: 0.003057071939110756
step: 170, loss: 0.0021129490341991186
step: 180, loss: 0.08367259800434113
step: 190, loss: 0.005480957217514515
step: 200, loss: 0.0018703287933021784
step: 210, loss: 0.0006624827510677278
step: 220, loss: 0.007710788864642382
step: 230, loss: 0.001298972754739225
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9728506787330317, f1=0.9718785151856018,
best_f1=0.9663677130044843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008108117035590112
step: 10, loss: 0.001795366988517344
step: 20, loss: 0.001357018481940031
step: 30, loss: 0.041632551699876785
step: 40, loss: 0.0006810532649978995
step: 50, loss: 0.0019283922156319022
step: 60, loss: 0.0005141349392943084
step: 70, loss: 0.00023611364304088056
step: 80, loss: 0.001353409606963396
step: 90, loss: 0.02468673139810562
step: 100, loss: 0.003960150759667158
step: 110, loss: 0.0006733532645739615
step: 120, loss: 0.04803701490163803
step: 130, loss: 0.044367801398038864
step: 140, loss: 0.001795533113181591
step: 150, loss: 0.0012497635325416923
step: 160, loss: 0.09839887917041779
step: 170, loss: 0.03289075195789337
step: 180, loss: 0.0021804107818752527
step: 190, loss: 0.0005318415933288634
step: 200, loss: 0.0030455333180725574
step: 210, loss: 0.015582337975502014
step: 220, loss: 0.006781741976737976
step: 230, loss: 0.003686592448502779
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9732739420935412, f1=0.9690949227373068,
best_f1=0.9663677130044843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004758129362016916
step: 10, loss: 0.0003136450832244009
step: 20, loss: 0.001664352254010737
step: 30, loss: 0.0017577573889866471
step: 40, loss: 0.0004287567862775177
step: 50, loss: 0.11561909317970276
step: 60, loss: 0.0077521707862615585
step: 70, loss: 0.006145139690488577
step: 80, loss: 0.002158487681299448
step: 90, loss: 0.0008053687051869929
step: 100, loss: 0.0004584263078868389
step: 110, loss: 0.01771986484527588
step: 120, loss: 0.000237567990552634
step: 130, loss: 0.0020243176259100437
step: 140, loss: 0.025861049070954323
step: 150, loss: 0.0077832848764956
step: 160, loss: 0.036900974810123444
step: 170, loss: 0.003382649039849639
step: 180, loss: 0.0008221673197112978
step: 190, loss: 0.06508806347846985
step: 200, loss: 0.002425732556730509
step: 210, loss: 0.002933315234258771
step: 220, loss: 0.00022647320292890072
step: 230, loss: 0.0007601164397783577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.972129319955407, f1=0.9665924276169264,
best_f1=0.9663677130044843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001663557137362659
step: 10, loss: 0.0009076888673007488
step: 20, loss: 0.000753495201934129
step: 30, loss: 0.0007035959279164672
step: 40, loss: 0.006803943309932947
step: 50, loss: 0.008834055624902248
step: 60, loss: 0.010663196444511414
step: 70, loss: 0.0005428753211162984
step: 80, loss: 7.661668496439233e-05
step: 90, loss: 0.00020708004012703896

```

```

step: 100, loss: 0.0008300643530674279
step: 110, loss: 0.000945316394791007
step: 120, loss: 0.00041269080247730017
step: 130, loss: 0.009422561153769493
step: 140, loss: 0.00020293994748499244
step: 150, loss: 0.017299365252256393
step: 160, loss: 0.00048396753845736384
step: 170, loss: 0.0004524473624769598
step: 180, loss: 0.0002728436083998531
step: 190, loss: 0.015193464234471321
step: 200, loss: 0.004518893081694841
step: 210, loss: 0.0002682850754354149
step: 220, loss: 0.000198586072656326
step: 230, loss: 0.025006389245390892
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9763779527559054, f1=0.9662162162162162,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02707047015428543
step: 10, loss: 0.005310899578034878
step: 20, loss: 0.0014123725704848766
step: 30, loss: 0.00032571700285188854
step: 40, loss: 0.005214311648160219
step: 50, loss: 0.0003829045163001865
step: 60, loss: 0.0004220415139570832
step: 70, loss: 0.0010638510575518012
step: 80, loss: 0.0008197652059607208
step: 90, loss: 0.0007338020368479192
step: 100, loss: 0.0003558063763193786
step: 110, loss: 0.00041578960372135043
step: 120, loss: 0.000140405201818794
step: 130, loss: 0.013755281455814838
step: 140, loss: 6.809696060372517e-05
step: 150, loss: 0.00016451565898023546
step: 160, loss: 0.00022455396538134664
step: 170, loss: 0.00032182305585592985
step: 180, loss: 0.00014365576498676091
step: 190, loss: 6.60653313389048e-05
step: 200, loss: 0.0015598421450704336
step: 210, loss: 8.641228487249464e-05

```

```

step: 220, loss: 0.002279485808685422
step: 230, loss: 0.003808404318988323
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9742441209406495, f1=0.9686800894854586,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018186283705290407
step: 10, loss: 0.00019018020248040557
step: 20, loss: 0.009301282465457916
step: 30, loss: 0.0031668576411902905
step: 40, loss: 0.00020377554756123573
step: 50, loss: 7.169471064116806e-05
step: 60, loss: 0.1135854572057724
step: 70, loss: 0.0004952525487169623
step: 80, loss: 0.025567082688212395
step: 90, loss: 0.04606395959854126
step: 100, loss: 0.015016305260360241
step: 110, loss: 7.576630014227703e-05
step: 120, loss: 0.06135334447026253
step: 130, loss: 0.00017302199557889253
step: 140, loss: 0.10241469740867615
step: 150, loss: 9.639697236707434e-05
step: 160, loss: 0.006972941569983959
step: 170, loss: 0.00012206931569380686
step: 180, loss: 0.00039518409175798297
step: 190, loss: 0.00011600843572523445
step: 200, loss: 0.00013646381557919085
step: 210, loss: 0.000283225643215701
step: 220, loss: 0.020647555589675903
step: 230, loss: 0.02342815324664116
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9705215419501134, f1=0.9625425652667423,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010575461201369762
step: 10, loss: 0.001691114972345531
step: 20, loss: 0.00031171904993243515
step: 30, loss: 0.00018701725639402866
step: 40, loss: 0.0002944133011624217
step: 50, loss: 0.004161070566624403

```



```

step: 60, loss: 0.021061787381768227
step: 70, loss: 0.00027399524697102606
step: 80, loss: 0.00012445887841749936
step: 90, loss: 0.0030324296094477177
step: 100, loss: 0.0031521767377853394
step: 110, loss: 0.020944848656654358
step: 120, loss: 0.028273379430174828
step: 130, loss: 0.0007334584952332079
step: 140, loss: 0.04041501507163048
step: 150, loss: 0.0011645004851743579
step: 160, loss: 9.096172288991511e-05
step: 170, loss: 0.00017573458899278194
step: 180, loss: 0.00012394486111588776
step: 190, loss: 0.0016163343098014593
step: 200, loss: 5.293158392305486e-05
step: 210, loss: 5.581723962677643e-05
step: 220, loss: 7.90217673056759e-05
step: 230, loss: 0.0016521656652912498
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9720670391061451, f1=0.9709821428571428,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.064561603125185e-05
step: 10, loss: 0.00034029828384518623
step: 20, loss: 3.1771698559168726e-05
step: 30, loss: 0.00011025947605958208
step: 40, loss: 0.01833779737353325
step: 50, loss: 0.000731802370864898
step: 60, loss: 6.153308640932664e-05
step: 70, loss: 6.366035813698545e-05
step: 80, loss: 0.0002753772714640945
step: 90, loss: 7.242929859785363e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 7.398451998597011e-05
step: 110, loss: 0.003496476449072361
step: 120, loss: 0.0001670924830250442
step: 130, loss: 5.442189649329521e-05
step: 140, loss: 5.071806663181633e-05
step: 150, loss: 4.133680704399012e-05
step: 160, loss: 6.050239971955307e-05
step: 170, loss: 0.0008209454244934022
step: 180, loss: 0.0006937173893675208
step: 190, loss: 7.350868690991774e-05
step: 200, loss: 6.723791739204898e-05

```

```

step: 210, loss: 6.836831016698852e-05
step: 220, loss: 0.0001531067246105522
step: 230, loss: 0.00018693138554226607
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9753363228699552, f1=0.9675977653631285,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00036638849996961653
step: 10, loss: 0.0011741381604224443
step: 20, loss: 5.46398623555433e-05
step: 30, loss: 0.0021139895543456078
step: 40, loss: 6.713491166010499e-05
step: 50, loss: 0.001775307347998023
step: 60, loss: 0.012600323185324669
step: 70, loss: 6.95478156558238e-05
step: 80, loss: 0.00044644164154306054
step: 90, loss: 5.684775533154607e-05
step: 100, loss: 6.318220403045416e-05
step: 110, loss: 0.00011497590458020568
step: 120, loss: 7.586555148009211e-05
step: 130, loss: 9.460405999561772e-05
step: 140, loss: 4.252968210494146e-05
step: 150, loss: 7.021505734883249e-05
step: 160, loss: 0.0011440104572102427
step: 170, loss: 5.247510125627741e-05
step: 180, loss: 0.0010804269695654511
step: 190, loss: 6.462735473178327e-05
step: 200, loss: 0.01130271703004837
step: 210, loss: 6.892051169415936e-05
step: 220, loss: 0.07137919217348099
step: 230, loss: 0.0003651570004876703
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9740698985343857, f1=0.9684684684684683,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.13700909120962e-05
step: 10, loss: 0.0002149922656826675
step: 20, loss: 0.0001456001918995753
step: 30, loss: 0.013224046677350998
step: 40, loss: 0.0001318317954428494

```

```

step: 50, loss: 5.796156619908288e-05
step: 60, loss: 8.125713793560863e-05
step: 70, loss: 6.254906475078315e-05
step: 80, loss: 7.445852679666132e-05
step: 90, loss: 3.6122633900959045e-05
step: 100, loss: 8.277515007648617e-05
step: 110, loss: 5.926016456214711e-05
step: 120, loss: 4.848219032282941e-05
step: 130, loss: 0.0001057574845617637
step: 140, loss: 0.029923545196652412
step: 150, loss: 0.00015907795750536025
step: 160, loss: 7.533492316724733e-05
step: 170, loss: 0.00012952151882927865
step: 180, loss: 0.0001497870107414201
step: 190, loss: 3.19802165904548e-05
step: 200, loss: 8.047121082199737e-05
step: 210, loss: 0.0005679032183252275
step: 220, loss: 0.0003004200989380479
step: 230, loss: 4.775649358634837e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.972972972972973, f1=0.9718785151856018,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002947675238829106
step: 10, loss: 2.8438020308385603e-05
step: 20, loss: 0.0002007526345551014
step: 30, loss: 9.508008224656805e-05
step: 40, loss: 3.268063665018417e-05
step: 50, loss: 0.00032138696406036615
step: 60, loss: 0.009130263701081276
step: 70, loss: 5.040228643338196e-05
step: 80, loss: 0.00024879362899810076
step: 90, loss: 0.0011496585793793201
step: 100, loss: 0.002480783499777317
step: 110, loss: 0.00015002379950601608
step: 120, loss: 4.979530785931274e-05
step: 130, loss: 6.252848106669262e-05
step: 140, loss: 5.398227949626744e-05
step: 150, loss: 6.538871093653142e-05
step: 160, loss: 7.331955566769466e-05
step: 170, loss: 0.0004013401921838522
step: 180, loss: 0.00014188114437274635
step: 190, loss: 5.345359386410564e-05
step: 200, loss: 4.9666836275719106e-05

```

```

step: 210, loss: 0.00010074328019982204
step: 220, loss: 0.0002628641959745437
step: 230, loss: 3.826052852673456e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.972129319955407, f1=0.9699666295884317,
best_f1=0.9662162162162162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.861266279476695e-05
step: 10, loss: 0.00036148057552054524
step: 20, loss: 6.357653910527006e-05
step: 30, loss: 0.00028618783107958734
step: 40, loss: 8.769375563133508e-05
step: 50, loss: 5.289969340083189e-05
step: 60, loss: 6.287880387390032e-05
step: 70, loss: 0.0010180568788200617
step: 80, loss: 0.0010085892863571644
step: 90, loss: 3.276635834481567e-05
step: 100, loss: 7.087972335284576e-05
step: 110, loss: 7.9778787039686e-05
step: 120, loss: 8.605813491158187e-05
step: 130, loss: 0.00010292658407706767
step: 140, loss: 0.012800971046090126
step: 150, loss: 0.00032534735510125756
step: 160, loss: 0.007558072917163372
step: 170, loss: 4.16466427850537e-05
step: 180, loss: 5.6482069339836016e-05
step: 190, loss: 0.0001347020734101534
step: 200, loss: 0.0004109767614863813
step: 210, loss: 0.018350806087255478
step: 220, loss: 0.011614671908318996
step: 230, loss: 7.156826177379116e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9743016759776536, f1=0.972129319955407,
best_f1=0.9662162162162162
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16

```

```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled               : True
opt_level             : 02
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 337.14it/s]
load_f1 = 0.972972972972973
real_f1 = 0.9686800894854586
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 400.25it/s]

```

### 3.3.2 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \

```

```

--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: ``optimizer.step()`` before ``lr_scheduler.step()``. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>",  
UserWarning)

step: 0, loss: 0.7862043380737305

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 10, loss: 0.4449467957019806

step: 20, loss: 0.5007806420326233

step: 30, loss: 0.44905075430870056

step: 40, loss: 0.3855626583099365

step: 50, loss: 0.2620943784713745

step: 60, loss: 0.278632789850235

step: 70, loss: 0.1469222754240036

step: 80, loss: 0.24435478448867798

step: 90, loss: 0.1026570051908493

step: 100, loss: 0.36836451292037964

step: 110, loss: 0.05335890129208565

step: 120, loss: 0.04813605546951294

step: 130, loss: 0.07420815527439117

step: 140, loss: 0.07958541810512543

step: 150, loss: 0.030665764585137367

step: 160, loss: 0.19550707936286926

step: 170, loss: 0.20182642340660095

step: 180, loss: 0.13218137621879578

step: 190, loss: 0.04841502755880356

step: 200, loss: 0.11926186084747314

step: 210, loss: 0.10380785912275314

step: 220, loss: 0.08743137866258621

step: 230, loss: 0.06611057370901108

step: 240, loss: 0.0888323187828064

step: 250, loss: 0.04718439653515816

step: 260, loss: 0.04806123673915863

step: 270, loss: 0.019457995891571045

step: 280, loss: 0.13018473982810974

step: 290, loss: 0.0748339593410492

step: 300, loss: 0.25200986862182617

step: 310, loss: 0.057858701795339584

step: 320, loss: 0.09482353925704956

step: 330, loss: 0.18837450444698334

step: 340, loss: 0.16143159568309784

step: 350, loss: 0.15022195875644684

step: 360, loss: 0.05538793280720711

step: 370, loss: 0.09062138944864273

step: 380, loss: 0.17303749918937683

step: 390, loss: 0.012306217104196548

```

step: 400, loss: 0.017616990953683853
step: 410, loss: 0.10271178185939789
step: 420, loss: 0.01229423750191927
step: 430, loss: 0.04942720755934715
step: 440, loss: 0.07614080607891083
step: 450, loss: 0.014247039332985878
step: 460, loss: 0.03150171786546707
step: 470, loss: 0.1901022046804428
step: 480, loss: 0.30346739292144775
step: 490, loss: 0.049315374344587326
step: 500, loss: 0.023773765191435814
step: 510, loss: 0.030440326780080795
step: 520, loss: 0.03859968110918999
step: 530, loss: 0.10282506793737411
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9344490934449092, f1=0.9274826789838336,
best_f1=0.9274826789838336
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10192415863275528
step: 10, loss: 0.1453043669462204
step: 20, loss: 0.06058230996131897
step: 30, loss: 0.08053265511989594
step: 40, loss: 0.009161200374364853
step: 50, loss: 0.04387933015823364
step: 60, loss: 0.1215069517493248
step: 70, loss: 0.14415156841278076
step: 80, loss: 0.009434572421014309
step: 90, loss: 0.053833670914173126
step: 100, loss: 0.20139175653457642
step: 110, loss: 0.06281302124261856
step: 120, loss: 0.07109596580266953
step: 130, loss: 0.0336422435939312
step: 140, loss: 0.04293671250343323
step: 150, loss: 0.05279209464788437
step: 160, loss: 0.008049539290368557
step: 170, loss: 0.13790835440158844
step: 180, loss: 0.00955837219953537
step: 190, loss: 0.025515377521514893
step: 200, loss: 0.020206039771437645
step: 210, loss: 0.006558115128427744

```



```

step: 220, loss: 0.09831856936216354
step: 230, loss: 0.038259901106357574
step: 240, loss: 0.12346374988555908
step: 250, loss: 0.04498710110783577
step: 260, loss: 0.027518656104803085
step: 270, loss: 0.17631813883781433
step: 280, loss: 0.1317417472600937
step: 290, loss: 0.10783474147319794
step: 300, loss: 0.05534495413303375
step: 310, loss: 0.035541269928216934
step: 320, loss: 0.23523397743701935
step: 330, loss: 0.033463094383478165
step: 340, loss: 0.009721513837575912
step: 350, loss: 0.028495434671640396
step: 360, loss: 0.02321789786219597
step: 370, loss: 0.009083723649382591
step: 380, loss: 0.07778111100196838
step: 390, loss: 0.032231803983449936
step: 400, loss: 0.04736503213644028
step: 410, loss: 0.00048579039867036045
step: 420, loss: 0.05934001877903938
step: 430, loss: 0.020273357629776
step: 440, loss: 0.06771189719438553
step: 450, loss: 0.019780360162258148
step: 460, loss: 0.22472701966762543
step: 470, loss: 0.02038213610649109
step: 480, loss: 0.21781949698925018
step: 490, loss: 0.03423050045967102
step: 500, loss: 0.022260520607233047
step: 510, loss: 0.1052931472659111
step: 520, loss: 0.0907544195652008
step: 530, loss: 0.2833021581172943
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9393939393939394, f1=0.9346826126954921,
best_f1=0.9346826126954921
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012101803906261921
step: 10, loss: 0.19184815883636475
step: 20, loss: 0.23574049770832062
step: 30, loss: 0.12048069387674332

```

step: 40, loss: 0.000611002673394978  
step: 50, loss: 0.020493166521191597  
step: 60, loss: 0.001765992259606719  
step: 70, loss: 0.032644122838974  
step: 80, loss: 0.009152263402938843  
step: 90, loss: 0.04431765526533127  
step: 100, loss: 0.03932959586381912  
step: 110, loss: 0.03741367906332016  
step: 120, loss: 0.0074810148216784  
step: 130, loss: 0.008652247488498688  
step: 140, loss: 0.10093769431114197  
step: 150, loss: 0.008832012303173542  
step: 160, loss: 0.0008921651751734316  
step: 170, loss: 0.006975094322115183  
step: 180, loss: 0.009057288058102131  
step: 190, loss: 0.07614966481924057  
step: 200, loss: 0.010841136798262596  
step: 210, loss: 0.12282353639602661  
step: 220, loss: 0.08497682213783264  
step: 230, loss: 0.03218579664826393  
step: 240, loss: 0.07283268123865128  
step: 250, loss: 0.011372005566954613  
step: 260, loss: 0.0005896901129744947  
step: 270, loss: 0.0014997647376731038  
step: 280, loss: 0.023373432457447052  
step: 290, loss: 0.046415288001298904  
step: 300, loss: 0.11578666418790817  
step: 310, loss: 0.10914085060358047  
step: 320, loss: 0.08418241143226624  
step: 330, loss: 0.020591145381331444  
step: 340, loss: 0.006231940351426601  
step: 350, loss: 0.019063621759414673  
step: 360, loss: 0.009063724428415298  
step: 370, loss: 0.0026842860970646143  
step: 380, loss: 0.0028006748761981726  
step: 390, loss: 0.0017664666520431638  
step: 400, loss: 0.022613754495978355  
step: 410, loss: 0.011036302894353867  
step: 420, loss: 0.0709562674164772  
step: 430, loss: 0.16284306347370148  
step: 440, loss: 0.013836199417710304  
step: 450, loss: 0.03520666062831879  
step: 460, loss: 0.060750171542167664  
step: 470, loss: 0.0016676642699167132  
step: 480, loss: 0.024358512833714485  
step: 490, loss: 0.0040949261747300625  
step: 500, loss: 0.07872865349054337  
step: 510, loss: 0.0016502587823197246

```

step: 520, loss: 0.0006144211511127651
step: 530, loss: 0.06927157193422318
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9399624765478424, f1=0.9361305361305362,
best_f1=0.9361305361305362
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001487893401645124
step: 10, loss: 0.0073658679611980915
step: 20, loss: 0.038041263818740845
step: 30, loss: 0.01813034899532795
step: 40, loss: 0.003334109438583255
step: 50, loss: 0.0011982036521658301
step: 60, loss: 0.001391862635500729
step: 70, loss: 0.0009387840982526541
step: 80, loss: 0.007694499101489782
step: 90, loss: 0.0005558711127378047
step: 100, loss: 0.0010293161030858755
step: 110, loss: 0.000195393426110968
step: 120, loss: 0.04462117329239845
step: 130, loss: 0.0055002253502607346
step: 140, loss: 0.02211429551243782
step: 150, loss: 0.0041649662889540195
step: 160, loss: 0.0023649237118661404
step: 170, loss: 0.006075491197407246
step: 180, loss: 0.0024458880070596933
step: 190, loss: 0.0024460717104375362
step: 200, loss: 0.0030851021874696016
step: 210, loss: 0.1468639373779297
step: 220, loss: 0.0025518692564219236
step: 230, loss: 0.21502099931240082
step: 240, loss: 0.0028067277744412422
step: 250, loss: 0.0038139834068715572
step: 260, loss: 0.07874646037817001
step: 270, loss: 0.002320407424122095
step: 280, loss: 0.0015316801145672798
step: 290, loss: 0.03111710026860237
step: 300, loss: 0.11553189158439636
step: 310, loss: 0.0027406555600464344
step: 320, loss: 0.08762474358081818
step: 330, loss: 0.043700210750103

```

```
step: 340, loss: 0.005012115463614464
step: 350, loss: 0.0020886706188321114
step: 360, loss: 0.006255720276385546
step: 370, loss: 0.029240494593977928
step: 380, loss: 0.0009930081432685256
step: 390, loss: 0.0374925434589386
step: 400, loss: 0.029574641957879066
step: 410, loss: 0.009736168198287487
step: 420, loss: 0.0005760209169238806
step: 430, loss: 0.00207883189432323
step: 440, loss: 0.03063349984586239
step: 450, loss: 0.014686414040625095
step: 460, loss: 0.0007260803249664605
step: 470, loss: 0.006610063835978508
step: 480, loss: 0.001898484886623919
step: 490, loss: 0.02088485099375248
step: 500, loss: 0.01308993436396122
step: 510, loss: 0.022296784445643425
step: 520, loss: 0.13677959144115448
step: 530, loss: 0.0005486493464559317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9394650398873767, f1=0.9304063521718824,
best_f1=0.9361305361305362
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021643415093421936
step: 10, loss: 0.0031910480465739965
step: 20, loss: 0.0019264599541202188
step: 30, loss: 0.05194811522960663
step: 40, loss: 0.0262765996158123
step: 50, loss: 0.002168954350054264
step: 60, loss: 0.038349561393260956
step: 70, loss: 0.0002265133080072701
step: 80, loss: 0.0005041997064836323
step: 90, loss: 0.04729384928941727
step: 100, loss: 0.0014048753073439002
step: 110, loss: 0.021382173523306847
step: 120, loss: 0.014380238018929958
step: 130, loss: 0.0014908153098076582
step: 140, loss: 0.013449267484247684
step: 150, loss: 0.0007677425164729357
step: 160, loss: 0.02275535650551319
step: 170, loss: 0.0923076719045639
step: 180, loss: 0.0010194925125688314
step: 190, loss: 0.00012877126573584974
```

```

step: 200, loss: 0.009585157968103886
step: 210, loss: 0.0057098702527582645
step: 220, loss: 4.850105688092299e-05
step: 230, loss: 0.00020790139387827367
step: 240, loss: 0.020130939781665802
step: 250, loss: 0.0006011511432006955
step: 260, loss: 0.0009705722914077342
step: 270, loss: 0.00831337459385395
step: 280, loss: 0.0035745869390666485
step: 290, loss: 0.00011196007835678756
step: 300, loss: 0.019347311928868294
step: 310, loss: 0.010364570654928684
step: 320, loss: 0.09598128497600555
step: 330, loss: 0.0006809425540268421
step: 340, loss: 0.004188159946352243
step: 350, loss: 0.0026077902875840664
step: 360, loss: 0.039906371384859085
step: 370, loss: 0.0008189105428755283
step: 380, loss: 0.13525812327861786
step: 390, loss: 0.0068418169394135475
step: 400, loss: 0.01455550454556942
step: 410, loss: 0.0002846504212357104
step: 420, loss: 0.010081115178763866
step: 430, loss: 0.0013440412003546953
step: 440, loss: 0.0007421577465720475
step: 450, loss: 0.008600332774221897
step: 460, loss: 0.0005520898848772049
step: 470, loss: 0.035267286002635956
step: 480, loss: 0.0025939117185771465
step: 490, loss: 0.016891757026314735
step: 500, loss: 0.0029461244121193886
step: 510, loss: 0.004169009625911713
step: 520, loss: 0.001188698341138661
step: 530, loss: 0.0319560170173645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9456928838951311, f1=0.9354988399071927,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007579278666526079
step: 10, loss: 0.00026208703638985753

```

step: 20, loss: 0.0005892952322028577  
step: 30, loss: 7.481329521397129e-05  
step: 40, loss: 0.001178547041490674  
step: 50, loss: 8.882817201083526e-05  
step: 60, loss: 5.391337981563993e-05  
step: 70, loss: 0.01470582839101553  
step: 80, loss: 0.0015649785054847598  
step: 90, loss: 0.0007236811798065901  
step: 100, loss: 0.024110835045576096  
step: 110, loss: 0.0016256527742370963  
step: 120, loss: 0.0010444815270602703  
step: 130, loss: 0.0007751218508929014  
step: 140, loss: 0.0022565375547856092  
step: 150, loss: 0.0033478657715022564  
step: 160, loss: 0.00013162520190235227  
step: 170, loss: 0.00014978293620515615  
step: 180, loss: 0.00023186425096355379  
step: 190, loss: 0.000631827802862972  
step: 200, loss: 8.656907448312268e-05  
step: 210, loss: 0.0065570189617574215  
step: 220, loss: 0.00035634503001347184  
step: 230, loss: 0.0002659503952600062  
step: 240, loss: 0.00045014804345555604  
step: 250, loss: 0.00041030559805221856  
step: 260, loss: 0.0004771637904923409  
step: 270, loss: 0.00014357078180182725  
step: 280, loss: 0.007220622152090073  
step: 290, loss: 0.0007087221601977944  
step: 300, loss: 0.0008261281764134765  
step: 310, loss: 0.00035107258008792996  
step: 320, loss: 0.000645303400233388  
step: 330, loss: 0.008127703331410885  
step: 340, loss: 0.00021667186229024082  
step: 350, loss: 0.023416876792907715  
step: 360, loss: 0.00013476480671670288  
step: 370, loss: 0.0067565906792879105  
step: 380, loss: 0.03720724582672119  
step: 390, loss: 0.031606823205947876  
step: 400, loss: 0.00024672510335221887  
step: 410, loss: 0.00045657571172341704  
step: 420, loss: 0.0033225149381905794  
step: 430, loss: 0.0011424691183492541  
step: 440, loss: 0.0006004118476994336  
step: 450, loss: 0.0002853689657058567  
step: 460, loss: 0.0002622113097459078  
step: 470, loss: 0.050878167152404785  
step: 480, loss: 0.06115856021642685  
step: 490, loss: 0.0003537749289534986

```

step: 500, loss: 0.002259396482259035
step: 510, loss: 0.00022365973563864827
step: 520, loss: 0.0012414309894666076
step: 530, loss: 0.0026044996920973063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9367924528301887, f1=0.9260480452190296,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001672315993346274
step: 10, loss: 0.00950980931520462
step: 20, loss: 0.006053518503904343
step: 30, loss: 0.012538011185824871
step: 40, loss: 0.0001982706307899207
step: 50, loss: 0.0013411527033895254
step: 60, loss: 0.01852431893348694
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.020230596885085106
step: 80, loss: 0.033933937549591064
step: 90, loss: 0.00018685267423279583
step: 100, loss: 0.1046205535531044
step: 110, loss: 0.00023026010603643954
step: 120, loss: 0.0003645842953119427
step: 130, loss: 0.0004170466272626072
step: 140, loss: 0.0006554509163834155
step: 150, loss: 0.00029738841112703085
step: 160, loss: 0.005045117810368538
step: 170, loss: 0.00022277396055869758
step: 180, loss: 0.00011377581540727988
step: 190, loss: 7.967217970872298e-05
step: 200, loss: 0.0005065834848210216
step: 210, loss: 0.0008112516952678561
step: 220, loss: 0.00018496988923288882
step: 230, loss: 0.00016517660696990788
step: 240, loss: 0.0007140490924939513
step: 250, loss: 0.00022949241974856704
step: 260, loss: 0.0026499524246901274
step: 270, loss: 5.280451659928076e-05
step: 280, loss: 0.011080699972808361
step: 290, loss: 0.0008771869470365345
step: 300, loss: 0.0007249640766531229
step: 310, loss: 7.35934081603773e-05
step: 320, loss: 0.032770294696092606
step: 330, loss: 0.00013760146975982934
step: 340, loss: 0.0006608542171306908

```

```

step: 350, loss: 0.0011958337854593992
step: 360, loss: 0.019154349341988564
step: 370, loss: 0.0024285248946398497
step: 380, loss: 0.00045263287029229105
step: 390, loss: 0.00018833669309969991
step: 400, loss: 0.00035896478220820427
step: 410, loss: 0.0001950428995769471
step: 420, loss: 0.00014514887880068272
step: 430, loss: 0.00019570882432162762
step: 440, loss: 0.0008928001625463367
step: 450, loss: 0.0005758037441410124
step: 460, loss: 0.0020721303299069405
step: 470, loss: 0.0026550665497779846
step: 480, loss: 0.01629248633980751
step: 490, loss: 0.0009879732970148325
step: 500, loss: 5.020244134357199e-05
step: 510, loss: 0.00026024747057817876
step: 520, loss: 0.00040579173946753144
step: 530, loss: 6.330156611511484e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9434315100514259, f1=0.9319029850746268,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004963736282661557
step: 10, loss: 0.0007954098982736468
step: 20, loss: 0.015079977922141552
step: 30, loss: 0.00041601245175115764
step: 40, loss: 0.0001635021180845797
step: 50, loss: 0.007699117064476013
step: 60, loss: 0.0006055901176296175
step: 70, loss: 0.0016499183839187026
step: 80, loss: 8.928724128054455e-05
step: 90, loss: 0.0004889259580522776
step: 100, loss: 0.00017553467478137463
step: 110, loss: 0.0004134407208766788
step: 120, loss: 0.0007169620366767049
step: 130, loss: 0.00018071131489705294
step: 140, loss: 3.51544949808158e-05
step: 150, loss: 0.0164532121270895
step: 160, loss: 0.0007258442929014564
step: 170, loss: 0.0001763466134434566
step: 180, loss: 7.648997416254133e-05
step: 190, loss: 0.00019740586867555976
step: 200, loss: 0.0004911083378829062

```



```

step: 210, loss: 0.00025780274881981313
step: 220, loss: 0.005923595745116472
step: 230, loss: 0.0009733471088111401
step: 240, loss: 0.0008810931467451155
step: 250, loss: 0.0038692972157150507
step: 260, loss: 0.0028748237527906895
step: 270, loss: 2.7186490115127526e-05
step: 280, loss: 0.0005410650628618896
step: 290, loss: 7.030979759292677e-05
step: 300, loss: 7.599325908813626e-05
step: 310, loss: 0.011668497696518898
step: 320, loss: 4.798291047336534e-05
step: 330, loss: 0.013791198842227459
step: 340, loss: 0.00010700344137148932
step: 350, loss: 0.00020468488219194114
step: 360, loss: 8.055942453211173e-05
step: 370, loss: 0.00016735473764128983
step: 380, loss: 0.000351091381162405
step: 390, loss: 7.138377259252593e-05
step: 400, loss: 0.03689167648553848
step: 410, loss: 0.0002611902018543333
step: 420, loss: 3.601086064008996e-05
step: 430, loss: 4.270669523975812e-05
step: 440, loss: 0.0019704645965248346
step: 450, loss: 5.032100671087392e-05
step: 460, loss: 0.00024396119988523424
step: 470, loss: 0.0001384141214657575
step: 480, loss: 0.0004136717470828444
step: 490, loss: 0.00022863774211145937
step: 500, loss: 0.001723280525766313
step: 510, loss: 0.00030623620841652155
step: 520, loss: 0.000491660030093044
step: 530, loss: 0.007430030032992363
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9410125406409661, f1=0.9345622119815669,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000270750664640218
step: 10, loss: 0.0003738777886610478
step: 20, loss: 0.0012917736312374473
step: 30, loss: 0.00011038079537684098
step: 40, loss: 0.00828488077968359
step: 50, loss: 0.0004646893939934671
step: 60, loss: 0.0005568982451222837

```

step: 70, loss: 0.051751505583524704  
step: 80, loss: 6.526338984258473e-05  
step: 90, loss: 0.0015488836215808988  
step: 100, loss: 0.0001894218585221097  
step: 110, loss: 0.0007289005443453789  
step: 120, loss: 4.1106584831140935e-05  
step: 130, loss: 7.198136154329404e-05  
step: 140, loss: 0.0016757226549088955  
step: 150, loss: 6.095719800214283e-05  
step: 160, loss: 8.849629375617951e-05  
step: 170, loss: 0.00011149272177135572  
step: 180, loss: 4.262046786607243e-05  
step: 190, loss: 0.00016115291509777308  
step: 200, loss: 0.0002212501858593896  
step: 210, loss: 3.446172922849655e-05  
step: 220, loss: 3.966530493926257e-05  
step: 230, loss: 3.7586312828352675e-05  
step: 240, loss: 0.0005279311444610357  
step: 250, loss: 0.00012680031068157405  
step: 260, loss: 7.833470590412617e-05  
step: 270, loss: 0.0002327934344066307  
step: 280, loss: 5.51807061128784e-05  
step: 290, loss: 2.979043529194314e-05  
step: 300, loss: 0.0004660379490815103  
step: 310, loss: 0.00012422996223904192  
step: 320, loss: 0.0008387499838136137  
step: 330, loss: 0.00011348970292601734  
step: 340, loss: 5.0343925977358595e-05  
step: 350, loss: 3.052423198823817e-05  
step: 360, loss: 0.013234877027571201  
step: 370, loss: 5.635123307001777e-05  
step: 380, loss: 2.602801214379724e-05  
step: 390, loss: 0.016182873398065567  
step: 400, loss: 0.00013370609667617828  
step: 410, loss: 6.542900518979877e-05  
step: 420, loss: 6.802817370044068e-05  
step: 430, loss: 2.615099037939217e-05  
step: 440, loss: 6.489529187092558e-05  
step: 450, loss: 3.6636993172578514e-05  
step: 460, loss: 5.8938789152307436e-05  
step: 470, loss: 4.682593134930357e-05  
step: 480, loss: 0.008123467676341534  
step: 490, loss: 0.00020958825189154595  
step: 500, loss: 0.0003417743428144604  
step: 510, loss: 0.0001858943869592622  
step: 520, loss: 0.020105600357055664  
step: 530, loss: 0.0007306004408746958  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9416705552963137, f1=0.9369786839666358,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00841111596673727
step: 10, loss: 3.188327536918223e-05
step: 20, loss: 0.0011267061345279217
step: 30, loss: 3.20515719067771e-05
step: 40, loss: 0.00023204066383186728
step: 50, loss: 3.0870651244185865e-05
step: 60, loss: 0.0005907385493628681
step: 70, loss: 3.765756991924718e-05
step: 80, loss: 0.0006435139803215861
step: 90, loss: 7.15549904271029e-05
step: 100, loss: 0.0001202272906084545
step: 110, loss: 0.0020087389275431633
step: 120, loss: 0.1340702623128891
step: 130, loss: 0.0003548681561369449
step: 140, loss: 0.007706908974796534
step: 150, loss: 0.00021863225265406072
step: 160, loss: 0.006705107633024454
step: 170, loss: 9.651482832850888e-05
step: 180, loss: 0.003800776321440935
step: 190, loss: 0.052539609372615814
step: 200, loss: 6.715852214256302e-05
step: 210, loss: 2.812533602991607e-05
step: 220, loss: 5.425952258519828e-05
step: 230, loss: 0.002933344803750515
step: 240, loss: 2.708604915824253e-05
step: 250, loss: 6.137423042673618e-05
step: 260, loss: 0.01846526935696602
step: 270, loss: 0.00012698175851255655
step: 280, loss: 3.286684295744635e-05
step: 290, loss: 3.799656769842841e-05
step: 300, loss: 3.22599480568897e-05
step: 310, loss: 0.0003975062572862953
step: 320, loss: 2.9182770958868787e-05
step: 330, loss: 0.00018598181486595422
step: 340, loss: 0.000167075137142092
step: 350, loss: 0.0002811262966133654
step: 360, loss: 3.7739348044851795e-05
step: 370, loss: 0.0010430788388475776
step: 380, loss: 4.0744347643340006e-05
step: 390, loss: 4.3123989598825574e-05
step: 400, loss: 4.483783050091006e-05
```

```

step: 410, loss: 6.514735287055373e-05
step: 420, loss: 0.00013573296018876135
step: 430, loss: 5.440115637611598e-05
step: 440, loss: 3.075879794778302e-05
step: 450, loss: 0.004835524130612612
step: 460, loss: 0.0004000840417575091
step: 470, loss: 0.0001067942357622087
step: 480, loss: 0.00011744978837668896
step: 490, loss: 0.015582149848341942
step: 500, loss: 0.00015215098392218351
step: 510, loss: 0.00042401449172757566
step: 520, loss: 4.549252116703428e-05
step: 530, loss: 3.046401252504438e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9379374708352777, f1=0.9314814814814815,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011313497816445306
step: 10, loss: 0.00022755799000151455
step: 20, loss: 5.9173522458877414e-05
step: 30, loss: 5.61519518669229e-05
step: 40, loss: 0.0039236824959516525
step: 50, loss: 0.000718506402336061
step: 60, loss: 5.958204565104097e-05
step: 70, loss: 3.808491237577982e-05
step: 80, loss: 0.00014206963533069938
step: 90, loss: 0.0005374650354497135
step: 100, loss: 5.4763007938163355e-05
step: 110, loss: 3.40254046022892e-05
step: 120, loss: 4.803669435204938e-05
step: 130, loss: 6.989040412008762e-05
step: 140, loss: 9.900127770379186e-05
step: 150, loss: 4.862451532972045e-05
step: 160, loss: 0.00014292853302322328
step: 170, loss: 0.002182723954319954
step: 180, loss: 4.1428440454183146e-05
step: 190, loss: 0.00016909194528125226
step: 200, loss: 0.00027317178319208324
step: 210, loss: 4.413050919538364e-05
step: 220, loss: 6.243753159651533e-05
step: 230, loss: 2.8002225008094683e-05
step: 240, loss: 2.5834313419181854e-05
step: 250, loss: 0.00024902037694118917
step: 260, loss: 2.6344432626501657e-05

```

```

step: 270, loss: 0.0007928143022581935
step: 280, loss: 4.262159927748144e-05
step: 290, loss: 0.0008067620801739395
step: 300, loss: 6.989934627199546e-05
step: 310, loss: 0.005430141929537058
step: 320, loss: 0.00799458660185337
step: 330, loss: 0.0014638375723734498
step: 340, loss: 4.1834973671939224e-05
step: 350, loss: 0.00492149218916893
step: 360, loss: 5.7763805671129376e-05
step: 370, loss: 2.7806925572804175e-05
step: 380, loss: 3.5432844015304e-05
step: 390, loss: 0.0021158072631806135
step: 400, loss: 2.1658350306097418e-05
step: 410, loss: 2.9175102099543437e-05
step: 420, loss: 6.508502701763064e-05
step: 430, loss: 3.378827386768535e-05
step: 440, loss: 3.2412575819762424e-05
step: 450, loss: 2.782342562568374e-05
step: 460, loss: 0.0002070491755148396
step: 470, loss: 0.0002571188088040799
step: 480, loss: 2.3416734620695934e-05
step: 490, loss: 0.003146382747218013
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 500, loss: 6.0391012084437534e-05
step: 510, loss: 2.348732050450053e-05
step: 520, loss: 1.98219386220444e-05
step: 530, loss: 1.7072745322366245e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9387947269303202, f1=0.9279700654817586,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.005263639963232e-05
step: 10, loss: 2.7506732294568792e-05
step: 20, loss: 2.8771650249836966e-05
step: 30, loss: 2.8110229322919622e-05
step: 40, loss: 8.547293691663072e-05
step: 50, loss: 3.386785101611167e-05
step: 60, loss: 0.00011191615340067074
step: 70, loss: 0.00011651065142359585
step: 80, loss: 0.007551555056124926
step: 90, loss: 0.025141840800642967
step: 100, loss: 0.00012936563871335238
step: 110, loss: 2.1077121346024796e-05

```

```

step: 120, loss: 0.0005344677483662963
step: 130, loss: 3.7213369068922475e-05
step: 140, loss: 5.359321221476421e-05
step: 150, loss: 0.00024039075651671737
step: 160, loss: 4.008843097835779e-05
step: 170, loss: 3.5261815355625004e-05
step: 180, loss: 0.00632176548242569
step: 190, loss: 3.0168450393830426e-05
step: 200, loss: 8.022809925023466e-05
step: 210, loss: 7.29099047021009e-05
step: 220, loss: 1.8775201169773936e-05
step: 230, loss: 0.001299518276937306
step: 240, loss: 5.2614697779063135e-05
step: 250, loss: 0.0006960539030842483
step: 260, loss: 3.475153062026948e-05
step: 270, loss: 4.493715096032247e-05
step: 280, loss: 2.0224237232469022e-05
step: 290, loss: 0.00026695537962950766
step: 300, loss: 7.748195639578626e-05
step: 310, loss: 4.270004137651995e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 320, loss: 0.07483763247728348
step: 330, loss: 1.7374501112499274e-05
step: 340, loss: 4.0080220060190186e-05
step: 350, loss: 0.005746173672378063
step: 360, loss: 0.00013569131260737777
step: 370, loss: 2.4552786271669902e-05
step: 380, loss: 2.4083406970021315e-05
step: 390, loss: 2.9003818781347945e-05
step: 400, loss: 0.00014503700367640704
step: 410, loss: 7.263612496899441e-05
step: 420, loss: 5.257956945570186e-05
step: 430, loss: 2.6634943424141966e-05
step: 440, loss: 1.858885480032768e-05
step: 450, loss: 0.014425322413444519
step: 460, loss: 8.359480125363916e-05
step: 470, loss: 3.1240666430676356e-05
step: 480, loss: 6.238844071049243e-05
step: 490, loss: 4.0317041566595435e-05
step: 500, loss: 0.00018032925436273217
step: 510, loss: 2.8448997909436002e-05
step: 520, loss: 0.0001970544399227947
step: 530, loss: 3.6581026506610215e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9407337723424272, f1=0.9307116104868914,
best_f1=0.9354988399071927

```

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:  
An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

step: 0, loss: 7.01165699865669e-05  
step: 10, loss: 1.7206892152898945e-05  
step: 20, loss: 0.014695260673761368  
step: 30, loss: 0.003703703870996833  
step: 40, loss: 1.941591835930012e-05  
step: 50, loss: 2.242193113488611e-05  
step: 60, loss: 3.100336471106857e-05  
step: 70, loss: 3.597932300181128e-05  
step: 80, loss: 1.8641103451955132e-05  
step: 90, loss: 5.112171857035719e-05  
step: 100, loss: 1.862594399426598e-05  
step: 110, loss: 2.8801479857065715e-05  
step: 120, loss: 0.006585316266864538  
step: 130, loss: 0.001721939304843545  
step: 140, loss: 0.002001530257984996  
step: 150, loss: 0.0005137532134540379  
step: 160, loss: 4.438230462255888e-05  
step: 170, loss: 5.0862323405453935e-05  
step: 180, loss: 0.0001503388921264559  
step: 190, loss: 6.25085667707026e-05  
step: 200, loss: 0.006665282417088747  
step: 210, loss: 0.000340520084137097  
step: 220, loss: 9.169911936623976e-05  
step: 230, loss: 1.671879545028787e-05  
step: 240, loss: 3.2702584576327354e-05  
step: 250, loss: 0.047319430857896805  
step: 260, loss: 1.758295729814563e-05  
step: 270, loss: 2.6812780561158434e-05  
step: 280, loss: 2.0111529011046514e-05  
step: 290, loss: 5.591191438725218e-05  
step: 300, loss: 5.2363342547323555e-05  
step: 310, loss: 3.3123647881438956e-05  
step: 320, loss: 2.784505886666011e-05  
step: 330, loss: 0.00015491976228076965  
step: 340, loss: 2.421725912427064e-05  
step: 350, loss: 0.018399061635136604  
step: 360, loss: 1.62904125318164e-05  
step: 370, loss: 1.846190934884362e-05  
step: 380, loss: 8.710029214853421e-05  
step: 390, loss: 1.9196097127860412e-05  
step: 400, loss: 1.878610419225879e-05  
step: 410, loss: 4.056996476720087e-05  
step: 420, loss: 2.2246755179367028e-05  
step: 430, loss: 0.008041842840611935  
step: 440, loss: 0.0001310770312556997

```

step: 450, loss: 4.638379323296249e-05
step: 460, loss: 1.714350401016418e-05
step: 470, loss: 0.04694179818034172
step: 480, loss: 1.7970318367588334e-05
step: 490, loss: 2.4116608983604237e-05
step: 500, loss: 2.4280954676214606e-05
step: 510, loss: 3.3015519875334576e-05
step: 520, loss: 2.045870314759668e-05
step: 530, loss: 2.3077178411767818e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9409559512652297, f1=0.9331476323119777,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.7592011974775232e-05
step: 10, loss: 3.951738108298741e-05
step: 20, loss: 1.737819184199907e-05
step: 30, loss: 2.622537431307137e-05
step: 40, loss: 1.7467598809162155e-05
step: 50, loss: 0.006265553180128336
step: 60, loss: 3.203184314770624e-05
step: 70, loss: 0.0034408799838274717
step: 80, loss: 5.9380472521297634e-05
step: 90, loss: 1.3738734196522273e-05
step: 100, loss: 3.955506326747127e-05
step: 110, loss: 1.7154618035419844e-05
step: 120, loss: 3.2130032195709646e-05
step: 130, loss: 2.477933230693452e-05
step: 140, loss: 0.0003736871003638953
step: 150, loss: 0.00017899330123327672
step: 160, loss: 0.0012737874640151858
step: 170, loss: 4.217444802634418e-05
step: 180, loss: 1.4737068340764381e-05
step: 190, loss: 2.4306908017024398e-05
step: 200, loss: 2.7439567929832265e-05
step: 210, loss: 0.00014164707681629807
step: 220, loss: 0.00027787694125436246
step: 230, loss: 0.00033023449941538274
step: 240, loss: 1.9024764696951024e-05
step: 250, loss: 1.9132825400447473e-05
step: 260, loss: 1.3157574358046986e-05
step: 270, loss: 0.0016121783992275596
step: 280, loss: 1.3619529454445e-05
step: 290, loss: 0.02149011194705963
step: 300, loss: 0.0003366958408150822

```



```

step: 310, loss: 3.975947038270533e-05
step: 320, loss: 2.957298966066446e-05
step: 330, loss: 2.616830897750333e-05
step: 340, loss: 6.33181189186871e-05
step: 350, loss: 4.4668799091596156e-05
step: 360, loss: 0.002470729872584343
step: 370, loss: 5.648049045703374e-05
step: 380, loss: 1.3746168406214565e-05
step: 390, loss: 8.100400737021118e-05
step: 400, loss: 2.2559892386198044e-05
step: 410, loss: 1.5511921446886845e-05
step: 420, loss: 1.4629043107561301e-05
step: 430, loss: 0.0014966585440561175
step: 440, loss: 1.4163383639242966e-05
step: 450, loss: 2.522238173696678e-05
step: 460, loss: 6.797759124310687e-05
step: 470, loss: 1.8905288015957922e-05
step: 480, loss: 1.7597876649233513e-05
step: 490, loss: 4.0814025851432234e-05
step: 500, loss: 3.459414801909588e-05
step: 510, loss: 1.608186721568927e-05
step: 520, loss: 2.841527202690486e-05
step: 530, loss: 1.7240396118722856e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9388915206063476, f1=0.9279321714554876,
best_f1=0.9354988399071927
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 1.651394268264994e-05
step: 10, loss: 2.7212337954551913e-05
step: 20, loss: 3.060308881686069e-05
step: 30, loss: 2.6303283448214643e-05
step: 40, loss: 1.7269974705413915e-05
step: 50, loss: 1.607058038644027e-05
step: 60, loss: 1.545597660879139e-05
step: 70, loss: 4.212384737911634e-05
step: 80, loss: 2.0645009499276057e-05
step: 90, loss: 1.6778320059529506e-05
step: 100, loss: 1.8111981262336485e-05
step: 110, loss: 0.004530441947281361
step: 120, loss: 0.003022338729351759
step: 130, loss: 3.2583222491666675e-05
step: 140, loss: 1.6417039660154842e-05
step: 150, loss: 2.706958730414044e-05
step: 160, loss: 2.6898651412921026e-05

```

```

step: 170, loss: 1.910657920234371e-05
step: 180, loss: 1.983998663490638e-05
step: 190, loss: 3.732896948349662e-05
step: 200, loss: 0.0005595558905042708
step: 210, loss: 1.6718791812309064e-05
step: 220, loss: 1.351893661194481e-05
step: 230, loss: 2.2589296349906363e-05
step: 240, loss: 2.2302896468318067e-05
step: 250, loss: 2.483582284185104e-05
step: 260, loss: 2.4224636945291422e-05
step: 270, loss: 3.204348104191013e-05
step: 280, loss: 2.1349265807657503e-05
step: 290, loss: 0.00026266230270266533
step: 300, loss: 0.00010881128400797024
step: 310, loss: 2.4258361008833162e-05
step: 320, loss: 2.911887531809043e-05
step: 330, loss: 0.0006122160702943802
step: 340, loss: 4.2386131099192426e-05
step: 350, loss: 3.068147270823829e-05
step: 360, loss: 2.020558895310387e-05
step: 370, loss: 1.468492791900644e-05
step: 380, loss: 1.6283029253827408e-05
step: 390, loss: 1.509841786173638e-05
step: 400, loss: 1.5824825823074207e-05
step: 410, loss: 7.026202365523204e-05
step: 420, loss: 6.073535405448638e-05
step: 430, loss: 1.9505201635183766e-05
step: 440, loss: 0.00021069854847155511
step: 450, loss: 3.146533708786592e-05
step: 460, loss: 1.3742447663389612e-05
step: 470, loss: 0.0002523555012885481
step: 480, loss: 0.0029340016189962626
step: 490, loss: 1.6208532542805187e-05
step: 500, loss: 4.665177402785048e-05
step: 510, loss: 3.0999432055978104e-05
step: 520, loss: 1.674490158620756e-05
step: 530, loss: 1.8592641936265863e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9430291801760075, f1=0.9305108145421077,
best_f1=0.9354988399071927
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True

```

```

opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                  : True
opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:23, 242.82it/s]
load_f1 = 0.9388523047977423
real_f1 = 0.9392369288742346
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 244.39it/s]

```

## 4 DITTO

### 4.1 DITTO STRUCTURED

#### 4.1.1 Beer - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Structured/Beer \
        --batch_size 32 \

```

```

--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly

```

because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8613265752792358
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.28, f1=0.2692307692307693, best_f1=0.2692307692307693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3763071894645691
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.3870967741935484, f1=0.3783783783783784,
best_f1=0.3783783783783784
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32628464698791504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.5, f1=0.35294117647058826, best_f1=0.35294117647058826
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3349679112434387
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.4905660377358491, f1=0.34285714285714286,
best_f1=0.35294117647058826
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2458707094192505
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.5217391304347825, f1=0.38596491228070184,
best_f1=0.38596491228070184
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.281213641166687
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.5142857142857143, f1=0.4, best_f1=0.38596491228070184
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1658339500427246
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.5, f1=0.4, best_f1=0.38596491228070184
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2708968222141266
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 8: dev\_f1=0.5294117647058824, f1=0.4, best\_f1=0.4

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.20152050256729126

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 9: dev\_f1=0.5833333333333334, f1=0.47058823529411764, best\_f1=0.47058823529411764

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.19503216445446014

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 10: dev\_f1=0.7142857142857143, f1=0.5000000000000001, best\_f1=0.5000000000000001

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.16159693896770477

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

epoch 11: dev\_f1=0.7096774193548386, f1=0.4615384615384615, best\_f1=0.5000000000000001

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.15307959914207458

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

epoch 12: dev\_f1=0.6923076923076924, f1=0.5000000000000001, best\_f1=0.5000000000000001

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06349329650402069
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7096774193548386, f1=0.4615384615384615,
best_f1=0.5000000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1655832976102829
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7096774193548386, f1=0.4615384615384615,
best_f1=0.5000000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13705489039421082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7096774193548386, f1=0.4615384615384615,
best_f1=0.5000000000000001
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```



```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
91it [00:00, 120670.78it/s]
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
load_f1 = 0.6
real_f1 = 0.6
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 251.38it/s]

```

#### 4.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/DBLP-ACM \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --dk general \
    --dk product \
    --summarize \
    --da entry_swap \
    --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Structured/DBLP-ACM \

```

```

--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0 /usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate", UserWarning)

step: 0, loss: 0.8060280680656433

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 10, loss: 0.4639892578125

step: 20, loss: 0.5920902490615845

```

step: 30, loss: 0.42335137724876404
step: 40, loss: 0.21294443309307098
step: 50, loss: 0.05412864685058594
step: 60, loss: 0.08685263246297836
step: 70, loss: 0.22906285524368286
step: 80, loss: 0.19242994487285614
step: 90, loss: 0.012601989321410656
step: 100, loss: 0.037781186401844025
step: 110, loss: 0.024409987032413483
step: 120, loss: 0.018152130767703056
step: 130, loss: 0.016297481954097748
step: 140, loss: 0.022648239508271217
step: 150, loss: 0.08483882993459702
step: 160, loss: 0.16885894536972046
step: 170, loss: 0.004433969501405954
step: 180, loss: 0.0080661466345191
step: 190, loss: 0.01198243722319603
step: 200, loss: 0.030257800593972206
step: 210, loss: 0.012993685901165009
step: 220, loss: 0.005902384873479605
step: 230, loss: 0.02138587459921837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9744160177975528, f1=0.9707865168539327,
best_f1=0.9707865168539327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011200226843357086
step: 10, loss: 0.0008265374344773591
step: 20, loss: 0.001969679491594434
step: 30, loss: 0.008007095195353031
step: 40, loss: 0.009803516790270805
step: 50, loss: 0.14083971083164215
step: 60, loss: 0.0069547733291983604
step: 70, loss: 0.008287091739475727
step: 80, loss: 0.012605424970388412
step: 90, loss: 0.004847986623644829
step: 100, loss: 0.05017471686005592
step: 110, loss: 0.07231047004461288
step: 120, loss: 0.043530650436878204
step: 130, loss: 0.013285146094858646
step: 140, loss: 0.0852249413728714

```

```

step: 150, loss: 0.009435622952878475
step: 160, loss: 0.010211710818111897
step: 170, loss: 0.036606766283512115
step: 180, loss: 0.004824396688491106
step: 190, loss: 0.03506370633840561
step: 200, loss: 0.07777132838964462
step: 210, loss: 0.0118023082613945
step: 220, loss: 0.0009391434141434729
step: 230, loss: 0.010725722648203373
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9865168539325843, f1=0.9775784753363228,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006996093317866325
step: 10, loss: 0.05551847442984581
step: 20, loss: 0.0013558515347540379
step: 30, loss: 0.020767001435160637
step: 40, loss: 0.0415276475250721
step: 50, loss: 0.003384690498933196
step: 60, loss: 0.032978713512420654
step: 70, loss: 0.0019856207072734833
step: 80, loss: 0.08639394491910934
step: 90, loss: 0.002055912744253874
step: 100, loss: 0.0009494887781329453
step: 110, loss: 0.005110612604767084
step: 120, loss: 0.0047334483824670315
step: 130, loss: 0.0014839454088360071
step: 140, loss: 0.0032459564972668886
step: 150, loss: 0.0011788303963840008
step: 160, loss: 0.001434434438124299
step: 170, loss: 0.0023739985190331936
step: 180, loss: 0.006081895437091589
step: 190, loss: 0.0032898555509746075
step: 200, loss: 0.0022612803149968386
step: 210, loss: 0.014819491654634476
step: 220, loss: 0.006814847234636545
step: 230, loss: 0.09198161214590073
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

epoch 3: dev_f1=0.9831649831649831, f1=0.9841986455981941,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017952651251107454
step: 10, loss: 0.0006363611319102347
step: 20, loss: 0.0005487733869813383
step: 30, loss: 0.0006172830471768975
step: 40, loss: 0.0067489054054021835
step: 50, loss: 0.0007966888369992375
step: 60, loss: 0.0005036424845457077
step: 70, loss: 0.03929653763771057
step: 80, loss: 0.21599555015563965
step: 90, loss: 0.0029206317849457264
step: 100, loss: 0.0015082268510013819
step: 110, loss: 0.09425464272499084
step: 120, loss: 0.13018733263015747
step: 130, loss: 0.0035475848708301783
step: 140, loss: 0.004433848429471254
step: 150, loss: 0.0013776541454717517
step: 160, loss: 0.0011474841739982367
step: 170, loss: 0.03842604532837868
step: 180, loss: 0.08488267660140991
step: 190, loss: 0.007581432815641165
step: 200, loss: 0.0006594632868655026
step: 210, loss: 0.00046082420158199966
step: 220, loss: 0.0009230998693965375
step: 230, loss: 0.0003459578438196331
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9853768278965129, f1=0.9886877828054299,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007314204704016447
step: 10, loss: 0.00046786630991846323
step: 20, loss: 0.0004305708862375468
step: 30, loss: 0.0004692620423156768
step: 40, loss: 0.0011358667397871614
step: 50, loss: 0.00036733373417519033
step: 60, loss: 0.0449756383895874
step: 70, loss: 0.0002477533707860857
step: 80, loss: 0.00036423985147848725
step: 90, loss: 0.16932526230812073
step: 100, loss: 0.002678977558389306

```

```

step: 110, loss: 0.040021270513534546
step: 120, loss: 0.08358508348464966
step: 130, loss: 0.07064352184534073
step: 140, loss: 0.027535250410437584
step: 150, loss: 0.0002860864333342761
step: 160, loss: 0.019504137337207794
step: 170, loss: 0.006843134760856628
step: 180, loss: 0.0007590712630189955
step: 190, loss: 0.0004743803874589503
step: 200, loss: 0.004862227942794561
step: 210, loss: 0.004481544252485037
step: 220, loss: 0.0006324294954538345
step: 230, loss: 0.00040842860471457243
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9809203142536477, f1=0.9841628959276018,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005798657657578588
step: 10, loss: 0.0004480849893298
step: 20, loss: 0.0004704854218289256
step: 30, loss: 0.000345958600519225
step: 40, loss: 0.0013109460705891252
step: 50, loss: 0.010898777283728123
step: 60, loss: 0.01625148206949234
step: 70, loss: 0.0010762476595118642
step: 80, loss: 0.005508952308446169
step: 90, loss: 0.00119034165982157
step: 100, loss: 0.0005305063677951694
step: 110, loss: 0.10565176606178284
step: 120, loss: 0.0007509501883760095
step: 130, loss: 0.0024578184820711613
step: 140, loss: 0.0011211660457774997
step: 150, loss: 0.001605997676961124
step: 160, loss: 0.0002787876292131841
step: 170, loss: 0.0005768958362750709
step: 180, loss: 0.00048651822726242244
step: 190, loss: 0.0007366390782408416
step: 200, loss: 0.000596173107624054
step: 210, loss: 0.0004298503918107599
step: 220, loss: 0.00035986912553198636
step: 230, loss: 0.0022618123330175877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

epoch 6: dev_f1=0.9831649831649831, f1=0.9853768278965129,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13118137419223785
step: 10, loss: 0.002345574088394642
step: 20, loss: 0.0010096719488501549
step: 30, loss: 0.0005714645958505571
step: 40, loss: 0.0006698240176774561
step: 50, loss: 0.0002361465449212119
step: 60, loss: 0.001096420339308679
step: 70, loss: 0.011336684226989746
step: 80, loss: 0.00014649798686150461
step: 90, loss: 0.0008826465927995741
step: 100, loss: 0.0023668049834668636
step: 110, loss: 0.0037969937548041344
step: 120, loss: 0.012896968983113766
step: 130, loss: 0.0009804521687328815
step: 140, loss: 0.0002283412468386814
step: 150, loss: 0.0008571609505452216
step: 160, loss: 0.00022861886827740818
step: 170, loss: 0.0036131227388978004
step: 180, loss: 0.00019812070240732282
step: 190, loss: 0.0006112075643613935
step: 200, loss: 0.000580708438064903
step: 210, loss: 0.0001354703272227198
step: 220, loss: 0.00042752461740747094
step: 230, loss: 0.027694109827280045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9822616407982262, f1=0.9701657458563535,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017075229436159134
step: 10, loss: 0.03877473250031471
step: 20, loss: 0.0013012212002649903
step: 30, loss: 0.019950630143284798
step: 40, loss: 0.00015156951849348843
step: 50, loss: 0.031051378697156906
step: 60, loss: 0.00040145363891497254
step: 70, loss: 0.0009186755050905049
step: 80, loss: 0.007513590157032013
step: 90, loss: 0.000658765435218811
step: 100, loss: 0.00017332256538793445

```

```

step: 110, loss: 0.0006829174235463142
step: 120, loss: 0.0002327971888007596
step: 130, loss: 0.0010272295912727714
step: 140, loss: 0.09969059377908707
step: 150, loss: 0.000406594859668985
step: 160, loss: 0.00018701977387536317
step: 170, loss: 0.0021556636784225702
step: 180, loss: 0.0006061794701963663
step: 190, loss: 0.005788660142570734
step: 200, loss: 0.0009033342939801514
step: 210, loss: 0.0001488551642978564
step: 220, loss: 0.00013837184815201908
step: 230, loss: 0.06496718525886536
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9841986455981941, f1=0.9830124575311437,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011289265239611268
step: 10, loss: 0.00043905689381062984
step: 20, loss: 0.0007896406459622085
step: 30, loss: 0.0004596808576025069
step: 40, loss: 0.00016214491915889084
step: 50, loss: 0.0018433231161907315
step: 60, loss: 0.0008629203075543046
step: 70, loss: 0.00039418498636223376
step: 80, loss: 0.0475274883210659
step: 90, loss: 0.21223624050617218
step: 100, loss: 0.006301168352365494
step: 110, loss: 0.0001786600478226319
step: 120, loss: 0.028835074976086617
step: 130, loss: 0.00018796652148012072
step: 140, loss: 0.0004437835013959557
step: 150, loss: 0.00034404988400638103
step: 160, loss: 0.0018215609015896916
step: 170, loss: 0.00026123246061615646
step: 180, loss: 0.0007499155472032726
step: 190, loss: 9.76036099018529e-05
step: 200, loss: 0.00016558417701162398
step: 210, loss: 0.000406558538088575
step: 220, loss: 0.04562262445688248
step: 230, loss: 0.02356954663991928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```



```

epoch 9: dev_f1=0.983050847457627, f1=0.9784335981838819,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000701524259056896
step: 10, loss: 0.00014103238936513662
step: 20, loss: 0.000281296408502385
step: 30, loss: 0.00026774039724841714
step: 40, loss: 0.0016276444075629115
step: 50, loss: 0.012392868287861347
step: 60, loss: 0.011876014992594719
step: 70, loss: 0.00024983141338452697
step: 80, loss: 0.0005378099158406258
step: 90, loss: 0.0001414807775290683
step: 100, loss: 0.00016146064444910735
step: 110, loss: 0.00032021102379076183
step: 120, loss: 0.03772946074604988
step: 130, loss: 0.0005356453475542367
step: 140, loss: 0.1482669711112976
step: 150, loss: 0.0004141150275245309
step: 160, loss: 0.001151695498265326
step: 170, loss: 0.0003306173312012106
step: 180, loss: 0.0005812824820168316
step: 190, loss: 0.001361509901471436
step: 200, loss: 0.00015795297804288566
step: 210, loss: 9.44078856264241e-05
step: 220, loss: 0.02458081766963005
step: 230, loss: 0.0003444364992901683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9843400447427293, f1=0.9832026875699889,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013711204519495368
step: 10, loss: 0.00805703829973936
step: 20, loss: 7.525856199208647e-05
step: 30, loss: 0.0006848914199508727
step: 40, loss: 0.0011931078042834997
step: 50, loss: 0.00019382941536605358
step: 60, loss: 0.0003304180281702429
step: 70, loss: 0.0030736993066966534
step: 80, loss: 0.0022565543185919523
step: 90, loss: 0.00017324226791970432
step: 100, loss: 8.47986011649482e-05

```

```

step: 110, loss: 6.860611756565049e-05
step: 120, loss: 0.0005648118676617742
step: 130, loss: 7.825774810044095e-05
step: 140, loss: 0.00018533658294472843
step: 150, loss: 7.396328146569431e-05
step: 160, loss: 0.00025929868570528924
step: 170, loss: 0.00425906153395772
step: 180, loss: 0.019206011667847633
step: 190, loss: 0.00019557653286028653
step: 200, loss: 0.00021127624495420605
step: 210, loss: 0.0003009691135957837
step: 220, loss: 0.0001890148560050875
step: 230, loss: 0.02653021737933159
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9820224719101124, f1=0.9853438556933484,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017004680121317506
step: 10, loss: 0.000496396329253912
step: 20, loss: 7.573951734229922e-05
step: 30, loss: 0.0002871475007850677
step: 40, loss: 8.860614616423845e-05
step: 50, loss: 0.0001105268020182848
step: 60, loss: 0.003953410312533379
step: 70, loss: 0.0006479381117969751
step: 80, loss: 0.0005985426832921803
step: 90, loss: 9.581221092958003e-05
step: 100, loss: 7.03655241522938e-05
step: 110, loss: 0.0002570931101217866
step: 120, loss: 0.0002430194290354848
step: 130, loss: 0.0017738756723701954
step: 140, loss: 0.00013234996004030108
step: 150, loss: 0.00016420094470959157
step: 160, loss: 0.016157779842615128
step: 170, loss: 0.006807274650782347
step: 180, loss: 0.0007109557627700269
step: 190, loss: 0.00011591101065278053
step: 200, loss: 0.0077389804646372795
step: 210, loss: 0.00042980111902579665
step: 220, loss: 0.00738421268761158
step: 230, loss: 0.0012346883304417133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

epoch 12: dev_f1=0.9854423292273236, f1=0.9821029082774049,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006014220882207155
step: 10, loss: 0.00039870833279564977
step: 20, loss: 0.00016135605983436108
step: 30, loss: 0.020892299711704254
step: 40, loss: 0.0010348440846428275
step: 50, loss: 6.828387267887592e-05
step: 60, loss: 0.0001496137265348807
step: 70, loss: 8.151600195560604e-05
step: 80, loss: 0.00010378691513324156
step: 90, loss: 7.235466182464734e-05
step: 100, loss: 0.00014035016647540033
step: 110, loss: 8.02924987510778e-05
step: 120, loss: 0.00014628932694904506
step: 130, loss: 0.00011862043902510777
step: 140, loss: 0.0001623731222935021
step: 150, loss: 0.02640415169298649
step: 160, loss: 9.462863090448081e-05
step: 170, loss: 8.691811672179028e-05
step: 180, loss: 0.00019933225121349096
step: 190, loss: 0.006254732143133879
step: 200, loss: 0.0001960218360181898
step: 210, loss: 0.00010601642134133726
step: 220, loss: 5.969237463432364e-05
step: 230, loss: 6.531315011670813e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9864864864864865, f1=0.9853768278965129,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.8880024880636483e-05
step: 10, loss: 5.1955612434539944e-05
step: 20, loss: 0.0007526857662014663
step: 30, loss: 0.0020671491511166096
step: 40, loss: 0.00010172489419346675
step: 50, loss: 0.00014657991414424032
step: 60, loss: 7.42948250262998e-05
step: 70, loss: 0.00011973965592915192
step: 80, loss: 0.0003189786511939019
step: 90, loss: 0.0003401260473765433
step: 100, loss: 0.0001392856502206996

```

```

step: 110, loss: 0.0005098309484310448
step: 120, loss: 0.0005645919009111822
step: 130, loss: 0.00027202616911381483
step: 140, loss: 0.00014411448501050472
step: 150, loss: 9.696795314084738e-05
step: 160, loss: 4.7278153942897916e-05
step: 170, loss: 6.528567610075697e-05
step: 180, loss: 0.00011890438327100128
step: 190, loss: 0.046462204307317734
step: 200, loss: 4.63222895632498e-05
step: 210, loss: 0.011242026463150978
step: 220, loss: 9.132206469075754e-05
step: 230, loss: 7.788809307385236e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9865168539325843, f1=0.9865168539325843,
best_f1=0.9775784753363228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00032617789111100137
step: 10, loss: 6.110888352850452e-05
step: 20, loss: 5.303418220137246e-05
step: 30, loss: 0.0043369620107114315
step: 40, loss: 7.376590656349435e-05
step: 50, loss: 7.473107689293101e-05
step: 60, loss: 5.093260551802814e-05
step: 70, loss: 9.353757195640355e-05
step: 80, loss: 0.0022437183652073145
step: 90, loss: 4.350164090283215e-05
step: 100, loss: 0.00011666338104987517
step: 110, loss: 7.03012483427301e-05
step: 120, loss: 0.00034675595816224813
step: 130, loss: 0.00014290791295934469
step: 140, loss: 9.157262684311718e-05
step: 150, loss: 0.00026471365708857775
step: 160, loss: 5.144689930602908e-05
step: 170, loss: 0.0008872323087416589
step: 180, loss: 0.00029719644226133823
step: 190, loss: 0.000324403983540833
step: 200, loss: 0.0006431857473216951
step: 210, loss: 0.00036143153556622565
step: 220, loss: 0.0051130857318639755
step: 230, loss: 0.00012286183482501656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```
epoch 15: dev_f1=0.9865168539325843, f1=0.9853768278965129,
best_f1=0.9775784753363228
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.
```

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
2473it [00:11, 211.43it/s]
load_f1 = 0.9865168539325843
real_f1 = 0.9865168539325843
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
4398it [00:18, 232.24it/s]
```

#### 4.1.3 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
```

```

!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/DBLP-GoogleScholar \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.797466516494751
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4502904415130615
step: 20, loss: 0.5015013217926025
step: 30, loss: 0.4388135075569153
step: 40, loss: 0.4011688828468323
step: 50, loss: 0.2329392284154892
step: 60, loss: 0.14376886188983917
step: 70, loss: 0.06746578961610794
step: 80, loss: 0.09560172259807587
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.11937680840492249
step: 100, loss: 0.3506026864051819
step: 110, loss: 0.04660902917385101
step: 120, loss: 0.07562604546546936
step: 130, loss: 0.04673928767442703
step: 140, loss: 0.3723881244659424
step: 150, loss: 0.04884191229939461
step: 160, loss: 0.12014854699373245
step: 170, loss: 0.18474256992340088
step: 180, loss: 0.13435570895671844
step: 190, loss: 0.039811696857213974
step: 200, loss: 0.15370991826057434
step: 210, loss: 0.09143940359354019
step: 220, loss: 0.09235524386167526
step: 230, loss: 0.1677207052707672
step: 240, loss: 0.1332414597272873
step: 250, loss: 0.05914381146430969
step: 260, loss: 0.017096150666475296

```

```

step: 270, loss: 0.009685952216386795
step: 280, loss: 0.10627508163452148
step: 290, loss: 0.13604560494422913
step: 300, loss: 0.03668569028377533
step: 310, loss: 0.10035428404808044
step: 320, loss: 0.051937494426965714
step: 330, loss: 0.0796305388212204
step: 340, loss: 0.2039656788110733
step: 350, loss: 0.08823467046022415
step: 360, loss: 0.1235702782869339
step: 370, loss: 0.20063233375549316
step: 380, loss: 0.1739678978919983
step: 390, loss: 0.017810411751270294
step: 400, loss: 0.009425442665815353
step: 410, loss: 0.05217372998595238
step: 420, loss: 0.0305609293282032
step: 430, loss: 0.08811533451080322
step: 440, loss: 0.10929787904024124
step: 450, loss: 0.023001259192824364
step: 460, loss: 0.09224680811166763
step: 470, loss: 0.11482252925634384
step: 480, loss: 0.28809112310409546
step: 490, loss: 0.12052111327648163
step: 500, loss: 0.05647113174200058
step: 510, loss: 0.10655290633440018
step: 520, loss: 0.045193057507276535
step: 530, loss: 0.2120518982410431
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9296435272045028, f1=0.9161592505854801,
best_f1=0.9161592505854801
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09943601489067078
step: 10, loss: 0.14295519888401031
step: 20, loss: 0.19150957465171814
step: 30, loss: 0.07124150544404984
step: 40, loss: 0.015378315933048725
step: 50, loss: 0.09381646662950516
step: 60, loss: 0.06327930092811584
step: 70, loss: 0.04788029566407204
step: 80, loss: 0.005360698793083429

```



```
step: 90, loss: 0.06326483190059662
step: 100, loss: 0.275562584400177
step: 110, loss: 0.01881304755806923
step: 120, loss: 0.047070007771253586
step: 130, loss: 0.03778209164738655
step: 140, loss: 0.03989206627011299
step: 150, loss: 0.040790997445583344
step: 160, loss: 0.06475001573562622
step: 170, loss: 0.042276591062545776
step: 180, loss: 0.01076599396765232
step: 190, loss: 0.10374350845813751
step: 200, loss: 0.022626569494605064
step: 210, loss: 0.023963110521435738
step: 220, loss: 0.2040873020887375
step: 230, loss: 0.027852768078446388
step: 240, loss: 0.10309017449617386
step: 250, loss: 0.022456487640738487
step: 260, loss: 0.009057535789906979
step: 270, loss: 0.16912877559661865
step: 280, loss: 0.2774563729763031
step: 290, loss: 0.0593353807926178
step: 300, loss: 0.03786243498325348
step: 310, loss: 0.08891277760267258
step: 320, loss: 0.1961817741394043
step: 330, loss: 0.022110888734459877
step: 340, loss: 0.08225244283676147
step: 350, loss: 0.028277212753891945
step: 360, loss: 0.06559930741786957
step: 370, loss: 0.015178125351667404
step: 380, loss: 0.06369725614786148
step: 390, loss: 0.049751367419958115
step: 400, loss: 0.03792620077729225
step: 410, loss: 0.0015426126774400473
step: 420, loss: 0.10242915898561478
step: 430, loss: 0.02991773933172226
step: 440, loss: 0.011669873259961605
step: 450, loss: 0.020733986049890518
step: 460, loss: 0.23413830995559692
step: 470, loss: 0.10180822759866714
step: 480, loss: 0.19789476692676544
step: 490, loss: 0.030131880193948746
step: 500, loss: 0.027703775092959404
step: 510, loss: 0.11087477952241898
step: 520, loss: 0.0262753963470459
step: 530, loss: 0.0555315762758255
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9315323707498836, f1=0.9290023201856149,
best_f1=0.9290023201856149
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0786830484867096
step: 10, loss: 0.05952493101358414
step: 20, loss: 0.14479994773864746
step: 30, loss: 0.12051055580377579
step: 40, loss: 0.04303522780537605
step: 50, loss: 0.016113828867673874
step: 60, loss: 0.0054179597645998
step: 70, loss: 0.013792064040899277
step: 80, loss: 0.08026203513145447
step: 90, loss: 0.053809069097042084
step: 100, loss: 0.023019950836896896
step: 110, loss: 0.036709170788526535
step: 120, loss: 0.03825325891375542
step: 130, loss: 0.049020182341337204
step: 140, loss: 0.0109739163890481
step: 150, loss: 0.05531887710094452
step: 160, loss: 0.0013815404381603003
step: 170, loss: 0.00976018887013197
step: 180, loss: 0.06241396442055702
step: 190, loss: 0.028810925781726837
step: 200, loss: 0.006168344523757696
step: 210, loss: 0.10091157257556915
step: 220, loss: 0.049952536821365356
step: 230, loss: 0.01977608911693096
step: 240, loss: 0.07880900800228119
step: 250, loss: 0.008149899542331696
step: 260, loss: 0.01565650850534439
step: 270, loss: 0.009589853696525097
step: 280, loss: 0.014313850551843643
step: 290, loss: 0.0311608724296093
step: 300, loss: 0.20617248117923737
step: 310, loss: 0.08294626325368881
step: 320, loss: 0.15420116484165192
step: 330, loss: 0.004807792138308287
step: 340, loss: 0.0026711965911090374
step: 350, loss: 0.013866562396287918
step: 360, loss: 0.014997946098446846
step: 370, loss: 0.0019838791340589523
step: 380, loss: 0.0475345142185688

```

```

step: 390, loss: 0.00507738720625639
step: 400, loss: 0.018785813823342323
step: 410, loss: 0.03457499295473099
step: 420, loss: 0.07088538259267807
step: 430, loss: 0.022296316921710968
step: 440, loss: 0.029592309147119522
step: 450, loss: 0.05546019971370697
step: 460, loss: 0.0908965989947319
step: 470, loss: 0.009597701951861382
step: 480, loss: 0.003064997959882021
step: 490, loss: 0.019166534766554832
step: 500, loss: 0.07779848575592041
step: 510, loss: 0.008585312403738499
step: 520, loss: 0.002831126330420375
step: 530, loss: 0.025529302656650543
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9244159413650939, f1=0.9308291342189647,
best_f1=0.9290023201856149
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003362074727192521
step: 10, loss: 0.0033503915183246136
step: 20, loss: 0.057357680052518845
step: 30, loss: 0.01648120768368244
step: 40, loss: 0.017620529979467392
step: 50, loss: 0.006745323538780212
step: 60, loss: 0.007716475985944271
step: 70, loss: 0.0008884461130946875
step: 80, loss: 0.008656638674438
step: 90, loss: 0.00793174747377634
step: 100, loss: 0.030148502439260483
step: 110, loss: 0.0007857177988626063
step: 120, loss: 0.020491475239396095
step: 130, loss: 0.015033875592052937
step: 140, loss: 0.032086681574583054
step: 150, loss: 0.00022932686260901392
step: 160, loss: 0.0003193359007127583
step: 170, loss: 0.0011939766118302941
step: 180, loss: 0.014833011664450169
step: 190, loss: 0.03849044442176819
step: 200, loss: 0.0016864701174199581
step: 210, loss: 0.21532604098320007
step: 220, loss: 0.011343340389430523
step: 230, loss: 0.20497630536556244
step: 240, loss: 0.003254200331866741

```

```

step: 250, loss: 0.004575703758746386
step: 260, loss: 0.08930055797100067
step: 270, loss: 0.01388306450098753
step: 280, loss: 0.000645902007818222
step: 290, loss: 0.056327641010284424
step: 300, loss: 0.00937309768050909
step: 310, loss: 0.033385373651981354
step: 320, loss: 0.09309989213943481
step: 330, loss: 0.023375418037176132
step: 340, loss: 0.06814875453710556
step: 350, loss: 0.0037954854778945446
step: 360, loss: 0.009876424446702003
step: 370, loss: 0.08945658057928085
step: 380, loss: 0.005372512154281139
step: 390, loss: 0.07327205687761307
step: 400, loss: 0.048610687255859375
step: 410, loss: 0.024603532627224922
step: 420, loss: 0.04626122862100601
step: 430, loss: 0.022249149158596992
step: 440, loss: 0.06022363156080246
step: 450, loss: 0.00869069341570139
step: 460, loss: 0.00042379344813525677
step: 470, loss: 0.0024154451675713062
step: 480, loss: 0.013154536485671997
step: 490, loss: 0.053651437163352966
step: 500, loss: 0.023239625617861748
step: 510, loss: 0.05497981235384941
step: 520, loss: 0.030176017433404922
step: 530, loss: 0.005150746554136276
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9382716049382716, f1=0.9315693430656934,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0034092040732502937
step: 10, loss: 0.0006747048464603722
step: 20, loss: 0.0009842318249866366
step: 30, loss: 0.0029576136730611324
step: 40, loss: 0.05818779021501541
step: 50, loss: 0.007146756164729595
step: 60, loss: 0.00663198484107852

```

step: 70, loss: 0.0008303291397169232  
step: 80, loss: 0.003190124174579978  
step: 90, loss: 0.012735229916870594  
step: 100, loss: 0.0007061413489282131  
step: 110, loss: 0.025341298431158066  
step: 120, loss: 0.004714697599411011  
step: 130, loss: 0.008785753510892391  
step: 140, loss: 0.008267572149634361  
step: 150, loss: 0.0012145934160798788  
step: 160, loss: 0.1590544730424881  
step: 170, loss: 0.02773870714008808  
step: 180, loss: 0.03874329477548599  
step: 190, loss: 0.0014647269854322076  
step: 200, loss: 0.013465923257172108  
step: 210, loss: 0.0008480099495500326  
step: 220, loss: 0.00036962254671379924  
step: 230, loss: 0.0015615790616720915  
step: 240, loss: 0.033571239560842514  
step: 250, loss: 0.010423948988318443  
step: 260, loss: 0.058761559426784515  
step: 270, loss: 0.005568488035351038  
step: 280, loss: 0.14122386276721954  
step: 290, loss: 0.001633916632272303  
step: 300, loss: 0.05022897571325302  
step: 310, loss: 0.019368572160601616  
step: 320, loss: 0.04447212070226669  
step: 330, loss: 0.0015617187600582838  
step: 340, loss: 0.11022472381591797  
step: 350, loss: 0.003271890804171562  
step: 360, loss: 0.024199802428483963  
step: 370, loss: 0.03208237513899803  
step: 380, loss: 0.04810486361384392  
step: 390, loss: 0.0062243142165243626  
step: 400, loss: 0.09694787114858627  
step: 410, loss: 0.0008953039650805295  
step: 420, loss: 0.006217027083039284  
step: 430, loss: 0.002832084195688367  
step: 440, loss: 0.014164011925458908  
step: 450, loss: 0.07556042075157166  
step: 460, loss: 0.0015206753741949797  
step: 470, loss: 0.037255387753248215  
step: 480, loss: 0.011316489428281784  
step: 490, loss: 0.009406325407326221  
step: 500, loss: 0.04051920399069786  
step: 510, loss: 0.08151217550039291  
step: 520, loss: 0.0019236415391787887  
step: 530, loss: 0.027022896334528923  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning:

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9364858599907279, f1=0.9314442413162706,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005359315546229482
step: 10, loss: 0.002579179359599948
step: 20, loss: 0.003417030442506075
step: 30, loss: 0.00678073987364769
step: 40, loss: 0.11904105544090271
step: 50, loss: 0.002472929423674941
step: 60, loss: 0.027577610686421394
step: 70, loss: 0.002752643311396241
step: 80, loss: 0.010953236371278763
step: 90, loss: 0.0014936000807210803
step: 100, loss: 0.024722836911678314
step: 110, loss: 0.01241878792643547
step: 120, loss: 0.009436003863811493
step: 130, loss: 0.0018698411295190454
step: 140, loss: 0.002009050454944372
step: 150, loss: 0.004537550266832113
step: 160, loss: 0.00223229150287807
step: 170, loss: 0.0034033432602882385
step: 180, loss: 0.000727220787666738
step: 190, loss: 0.0019388777436688542
step: 200, loss: 9.979336755350232e-05
step: 210, loss: 0.0003910971281584352
step: 220, loss: 0.002116878516972065
step: 230, loss: 0.0002624689368531108
step: 240, loss: 0.004052694886922836
step: 250, loss: 0.009880868718028069
step: 260, loss: 0.008997159078717232
step: 270, loss: 0.0023712823167443275
step: 280, loss: 0.011748261749744415
step: 290, loss: 0.002606867579743266
step: 300, loss: 0.0011805560206994414
step: 310, loss: 0.0012657330371439457
step: 320, loss: 0.00369263649918139
step: 330, loss: 0.011315937153995037
step: 340, loss: 0.004377095028758049
step: 350, loss: 0.07214255630970001
step: 360, loss: 0.003915715962648392
step: 370, loss: 0.0022094575688242912
step: 380, loss: 0.04854077100753784
step: 390, loss: 0.02431105077266693
step: 400, loss: 0.007627092767506838
```

```

step: 410, loss: 0.0005868156440556049
step: 420, loss: 0.007697076536715031
step: 430, loss: 0.0001885399833554402
step: 440, loss: 0.08721448481082916
step: 450, loss: 0.0053899409249424934
step: 460, loss: 0.001238396274857223
step: 470, loss: 0.19178803265094757
step: 480, loss: 0.038158465176820755
step: 490, loss: 0.0006588235846720636
step: 500, loss: 0.00498005049303174
step: 510, loss: 0.006580775603652
step: 520, loss: 0.0008251852123066783
step: 530, loss: 0.003860775614157319
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9304911955514366, f1=0.9204597701149425,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007865781895816326
step: 10, loss: 0.035983119159936905
step: 20, loss: 0.19168777763843536
step: 30, loss: 0.019845718517899513
step: 40, loss: 0.022760361433029175
step: 50, loss: 0.0027656247839331627
step: 60, loss: 0.009005045518279076
step: 70, loss: 0.004103874322026968
step: 80, loss: 0.04508965462446213
step: 90, loss: 0.00033373170299455523
step: 100, loss: 0.008762157522141933
step: 110, loss: 0.01597774028778076
step: 120, loss: 0.002531748265028
step: 130, loss: 0.016059283167123795
step: 140, loss: 0.007061845622956753
step: 150, loss: 0.00042914925143122673
step: 160, loss: 0.0007529876893386245
step: 170, loss: 0.001711123390123248
step: 180, loss: 0.00029911179444752634
step: 190, loss: 0.0009261510567739606
step: 200, loss: 0.0022169346921145916
step: 210, loss: 0.00013621774269267917
step: 220, loss: 0.004436125047504902
step: 230, loss: 0.013593098148703575
step: 240, loss: 0.09711664915084839
step: 250, loss: 0.00022862311743665487
step: 260, loss: 0.0006243673851713538

```

```

step: 270, loss: 0.001903287135064602
step: 280, loss: 0.011793294921517372
step: 290, loss: 0.005680285859853029
step: 300, loss: 0.06053939089179039
step: 310, loss: 0.001321829971857369
step: 320, loss: 0.027291733771562576
step: 330, loss: 0.004208199214190245
step: 340, loss: 0.006306154653429985
step: 350, loss: 0.0003100415924564004
step: 360, loss: 0.0037013068795204163
step: 370, loss: 0.0004816541913896799
step: 380, loss: 0.0006785988807678223
step: 390, loss: 0.0006721084355376661
step: 400, loss: 0.00031881380709819496
step: 410, loss: 0.01976093277335167
step: 420, loss: 0.001172918826341629
step: 430, loss: 0.0003715587663464248
step: 440, loss: 0.00018005473248194903
step: 450, loss: 0.0006299674278125167
step: 460, loss: 0.0020677654538303614
step: 470, loss: 0.05775671824812889
step: 480, loss: 0.004474463872611523
step: 490, loss: 0.0031686481088399887
step: 500, loss: 0.0002788263955153525
step: 510, loss: 0.005212484858930111
step: 520, loss: 0.0005806381232105196
step: 530, loss: 0.0031982071232050657
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9334557136301056, f1=0.9298325033952014,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004140831995755434
step: 10, loss: 0.01208678912371397
step: 20, loss: 0.016910362988710403
step: 30, loss: 0.012108239345252514
step: 40, loss: 0.0035779743921011686
step: 50, loss: 0.0002198712172685191
step: 60, loss: 7.364709745161235e-05
step: 70, loss: 0.00011175268446095288
step: 80, loss: 0.020795080810785294
step: 90, loss: 0.00015249045100063086
step: 100, loss: 0.01841983012855053
step: 110, loss: 0.002087669912725687
step: 120, loss: 0.005616025999188423

```



```
step: 130, loss: 0.000146782593219541
step: 140, loss: 5.0403945351717994e-05
step: 150, loss: 0.0003682599635794759
step: 160, loss: 0.005281833931803703
step: 170, loss: 0.018856024369597435
step: 180, loss: 0.0003407225012779236
step: 190, loss: 0.0023023351095616817
step: 200, loss: 0.0009079679730348289
step: 210, loss: 0.00608806312084198
step: 220, loss: 0.0024981049355119467
step: 230, loss: 0.003319674637168646
step: 240, loss: 0.0019609197042882442
step: 250, loss: 0.00012399550178088248
step: 260, loss: 0.06049206852912903
step: 270, loss: 0.00015187951794359833
step: 280, loss: 0.006836137734353542
step: 290, loss: 0.10943292081356049
step: 300, loss: 0.00022399694717023522
step: 310, loss: 0.02308131754398346
step: 320, loss: 0.000589187431614846
step: 330, loss: 0.018796944990754128
step: 340, loss: 0.0011537086684256792
step: 350, loss: 0.007998250424861908
step: 360, loss: 0.0024099182337522507
step: 370, loss: 0.00015626476670149714
step: 380, loss: 0.0015226855175569654
step: 390, loss: 5.88059592701029e-05
step: 400, loss: 0.06698468327522278
step: 410, loss: 0.005189857445657253
step: 420, loss: 7.56393201299943e-05
step: 430, loss: 5.913557833991945e-05
step: 440, loss: 0.06491003185510635
step: 450, loss: 0.0009123377967625856
step: 460, loss: 0.0049571567215025425
step: 470, loss: 7.980110967764631e-05
step: 480, loss: 0.015204379335045815
step: 490, loss: 0.0007987680728547275
step: 500, loss: 0.0013998771319165826
step: 510, loss: 0.0003732214099727571
step: 520, loss: 0.0004283575981389731
step: 530, loss: 0.00013240473344922066
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9327188940092166, f1=0.9341262580054894,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006306131253950298
step: 10, loss: 0.032730210572481155
step: 20, loss: 0.0002458226808812469
step: 30, loss: 0.0002032016491284594
step: 40, loss: 0.016584916040301323
step: 50, loss: 0.00022068082762416452
step: 60, loss: 0.00040680525125935674
step: 70, loss: 0.11853053420782089
step: 80, loss: 0.00024860844132490456
step: 90, loss: 0.0017840625951066613
step: 100, loss: 0.00047294891555793583
step: 110, loss: 0.0021725327242165804
step: 120, loss: 0.0012710256269201636
step: 130, loss: 0.012304693460464478
step: 140, loss: 0.00326418224722147
step: 150, loss: 0.0001839998149080202
step: 160, loss: 0.03902148827910423
step: 170, loss: 0.0001853488211054355
step: 180, loss: 4.65560078737326e-05
step: 190, loss: 0.00025638804072514176
step: 200, loss: 0.0008452406036667526
step: 210, loss: 0.0001348217047052458
step: 220, loss: 0.00013126923295203596
step: 230, loss: 0.0007268652552738786
step: 240, loss: 0.0005787556292489171
step: 250, loss: 0.02634178288280964
step: 260, loss: 0.0002083940926240757
step: 270, loss: 0.006930581294000149
step: 280, loss: 0.0001584159763297066
step: 290, loss: 0.0002632123068906367
step: 300, loss: 0.000278785708360374
step: 310, loss: 4.1948991565732285e-05
step: 320, loss: 0.0008636745042167604
step: 330, loss: 0.0005052114138379693
step: 340, loss: 0.00028146724798716605
step: 350, loss: 5.173706085770391e-05
step: 360, loss: 0.019178122282028198
step: 370, loss: 0.00029408218688331544
step: 380, loss: 0.000317074591293931
step: 390, loss: 9.01747407624498e-05
step: 400, loss: 0.0005618929280899465
step: 410, loss: 0.0002819581422954798
step: 420, loss: 0.008250558748841286
step: 430, loss: 3.229708090657368e-05
step: 440, loss: 8.202929166145623e-05
step: 450, loss: 0.018010301515460014
step: 460, loss: 0.0052604759112000465
```

```

step: 470, loss: 8.601594890933484e-05
step: 480, loss: 0.0009976431028917432
step: 490, loss: 0.012592166662216187
step: 500, loss: 0.002171638421714306
step: 510, loss: 0.00015916794654913247
step: 520, loss: 5.001463068765588e-05
step: 530, loss: 0.017333364114165306
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9349555451567618, f1=0.9309865678554886,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004885037778876722
step: 10, loss: 0.00024455992388539016
step: 20, loss: 0.0004674948868341744
step: 30, loss: 3.531093170749955e-05
step: 40, loss: 0.0017182034207507968
step: 50, loss: 7.282268779817969e-05
step: 60, loss: 0.0012420928105711937
step: 70, loss: 0.00041768403025344014
step: 80, loss: 0.0005652903346344829
step: 90, loss: 0.0003035081608686596
step: 100, loss: 0.00022115797037258744
step: 110, loss: 0.0008444669074378908
step: 120, loss: 0.0005414090701378882
step: 130, loss: 0.0002768017293419689
step: 140, loss: 0.0002423139667371288
step: 150, loss: 0.006718150805681944
step: 160, loss: 0.00037168452399782836
step: 170, loss: 0.00015990028623491526
step: 180, loss: 0.06914880871772766
step: 190, loss: 0.00010530672443564981
step: 200, loss: 9.548074740450829e-05
step: 210, loss: 9.547016816213727e-05
step: 220, loss: 0.0009122189949266613
step: 230, loss: 0.00011015722702722996
step: 240, loss: 0.0001728327333694324
step: 250, loss: 0.00034350468195043504
step: 260, loss: 0.005888588726520538
step: 270, loss: 0.022546006366610527
step: 280, loss: 4.281720975995995e-05
step: 290, loss: 5.345067984308116e-05
step: 300, loss: 0.001992221223190427
step: 310, loss: 0.0021610124967992306
step: 320, loss: 0.002849593758583069

```

```

step: 330, loss: 0.010698283091187477
step: 340, loss: 0.0018128789961338043
step: 350, loss: 0.00021369790192693472
step: 360, loss: 0.0005484637804329395
step: 370, loss: 0.0013990085572004318
step: 380, loss: 0.0009101918549276888
step: 390, loss: 0.0005431552999652922
step: 400, loss: 0.0005305737722665071
step: 410, loss: 0.00012236730253789574
step: 420, loss: 0.0009348587482236326
step: 430, loss: 6.343994027702138e-05
step: 440, loss: 4.4846441596746445e-05
step: 450, loss: 0.002491738647222519
step: 460, loss: 0.006600539665669203
step: 470, loss: 9.39199817366898e-05
step: 480, loss: 0.00015411156346090138
step: 490, loss: 0.006372937001287937
step: 500, loss: 0.03638504073023796
step: 510, loss: 0.0005545064341276884
step: 520, loss: 0.0007283187587745488
step: 530, loss: 8.610317308921367e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9328493647912885, f1=0.9282511210762331,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022038126189727336
step: 10, loss: 0.0023815548047423363
step: 20, loss: 0.002557475818321109
step: 30, loss: 0.03408702090382576
step: 40, loss: 0.0008729007095098495
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.000715973146725446
step: 60, loss: 0.1467135101556778
step: 70, loss: 0.00018184677173849195
step: 80, loss: 0.003730909898877144
step: 90, loss: 0.002087387954816222
step: 100, loss: 9.020633297041059e-05
step: 110, loss: 0.00031270869658328593
step: 120, loss: 0.04018559679389
step: 130, loss: 0.07676549255847931
step: 140, loss: 0.0001074809770216234
step: 150, loss: 0.0012335091596469283
step: 160, loss: 0.023079438135027885
step: 170, loss: 0.0005692822160199285

```

```

step: 180, loss: 0.00023372945724986494
step: 190, loss: 0.0006187529652379453
step: 200, loss: 0.006756870541721582
step: 210, loss: 0.000109286600491032
step: 220, loss: 0.0015210662968456745
step: 230, loss: 0.002626561326906085
step: 240, loss: 5.327122926246375e-05
step: 250, loss: 0.0003037111891899258
step: 260, loss: 0.00026344601064920425
step: 270, loss: 0.04045940935611725
step: 280, loss: 0.00022757804254069924
step: 290, loss: 0.002787642180919647
step: 300, loss: 0.048770543187856674
step: 310, loss: 0.0008498226525261998
step: 320, loss: 0.010993611998856068
step: 330, loss: 0.0014415154000744224
step: 340, loss: 0.004109491128474474
step: 350, loss: 0.002965358318760991
step: 360, loss: 0.0003233743191231042
step: 370, loss: 0.00011275963333901018
step: 380, loss: 3.158452091156505e-05
step: 390, loss: 0.01250496320426464
step: 400, loss: 4.7167399316094816e-05
step: 410, loss: 0.0012534980196505785
step: 420, loss: 0.0007246563327498734
step: 430, loss: 0.0010941323125734925
step: 440, loss: 0.0006343798013404012
step: 450, loss: 0.00011053761409129947
step: 460, loss: 0.07476960122585297
step: 470, loss: 0.0006688260473310947
step: 480, loss: 0.0002179117000196129
step: 490, loss: 0.026258906349539757
step: 500, loss: 0.0004940733197145164
step: 510, loss: 5.634712942992337e-05
step: 520, loss: 2.4801989638945088e-05
step: 530, loss: 2.7044579837820493e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9376744186046512, f1=0.9340101522842639,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001508056651800871
step: 10, loss: 0.00016805522318463773
step: 20, loss: 0.00014605092292185873
step: 30, loss: 0.0001672526850597933

```

step: 40, loss: 6.366775778587908e-05  
step: 50, loss: 0.00609787879511714  
step: 60, loss: 0.0003198522317688912  
step: 70, loss: 0.0028315745294094086  
step: 80, loss: 0.0019011478871107101  
step: 90, loss: 0.0006373699870891869  
step: 100, loss: 0.0002987936313729733  
step: 110, loss: 0.00012191337009426206  
step: 120, loss: 0.017571212723851204  
step: 130, loss: 0.00015287288988474756  
step: 140, loss: 4.008215182693675e-05  
step: 150, loss: 0.00017418430070392787  
step: 160, loss: 0.00033883535070344806  
step: 170, loss: 0.0010789582738652825  
step: 180, loss: 0.023203076794743538  
step: 190, loss: 5.539720950764604e-05  
step: 200, loss: 0.00013095923350192606  
step: 210, loss: 0.00033765731495805085  
step: 220, loss: 3.1173745810519904e-05  
step: 230, loss: 0.0009559563477523625  
step: 240, loss: 0.0006175012094900012  
step: 250, loss: 9.277164645027369e-05  
step: 260, loss: 5.436045103124343e-05  
step: 270, loss: 0.0002392386959400028  
step: 280, loss: 0.0002757833572104573  
step: 290, loss: 0.020836884155869484  
step: 300, loss: 0.18369507789611816  
step: 310, loss: 6.188903353177011e-05  
step: 320, loss: 0.00017218371795024723  
step: 330, loss: 7.621980330441147e-05  
step: 340, loss: 0.0004994270857423544  
step: 350, loss: 0.016429966315627098  
step: 360, loss: 0.001999802188947797  
step: 370, loss: 0.0007557090721093118  
step: 380, loss: 7.201087282737717e-05  
step: 390, loss: 7.920418283902109e-05  
step: 400, loss: 2.9346741939662024e-05  
step: 410, loss: 0.008743392303586006  
step: 420, loss: 0.36498600244522095  
step: 430, loss: 0.010678904131054878  
step: 440, loss: 0.0048324549570679665  
step: 450, loss: 0.0012004534946754575  
step: 460, loss: 0.0011967234313488007  
step: 470, loss: 9.426313044968992e-05  
step: 480, loss: 0.0012098881416022778  
step: 490, loss: 0.00016800356388557702  
step: 500, loss: 0.0029026460833847523  
step: 510, loss: 6.967208901187405e-05

```
step: 520, loss: 0.019558968022465706
step: 530, loss: 0.00021884654415771365
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.934640522875817, f1=0.9332719742291763,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.98185680853203e-05
step: 10, loss: 0.00022478774189949036
step: 20, loss: 0.0310721043497324
step: 30, loss: 0.000194415872101672
step: 40, loss: 0.00014522875426337123
step: 50, loss: 0.0032120782416313887
step: 60, loss: 0.0008057424565777183
step: 70, loss: 0.00011565761815290898
step: 80, loss: 0.00011178322893101722
step: 90, loss: 0.00022216585057321936
step: 100, loss: 3.664941687020473e-05
step: 110, loss: 3.142178320558742e-05
step: 120, loss: 0.0018765456043183804
step: 130, loss: 0.002771081170067191
step: 140, loss: 2.6254985641571693e-05
step: 150, loss: 9.929219959303737e-05
step: 160, loss: 5.069784674560651e-05
step: 170, loss: 0.0014901498798280954
step: 180, loss: 0.00018694231403060257
step: 190, loss: 0.00025881006149575114
step: 200, loss: 0.002782366704195738
step: 210, loss: 0.002248141448944807
step: 220, loss: 4.0787628677207977e-05
step: 230, loss: 0.00046969743561930954
step: 240, loss: 3.2362349884351715e-05
step: 250, loss: 0.061902329325675964
step: 260, loss: 0.2613203227519989
step: 270, loss: 0.0032585004810243845
step: 280, loss: 0.0002704791259020567
step: 290, loss: 0.02377694472670555
step: 300, loss: 0.0008356343023478985
step: 310, loss: 9.000275895232335e-05
step: 320, loss: 0.0002518862020224333
step: 330, loss: 0.05908479541540146
step: 340, loss: 5.611268716165796e-05
step: 350, loss: 0.029539961367845535
step: 360, loss: 9.569516987539828e-05
step: 370, loss: 0.0015499066794291139
```

```

step: 380, loss: 0.0013605236308649182
step: 390, loss: 0.0008786341059021652
step: 400, loss: 0.00012171484559075907
step: 410, loss: 0.0008936085505411029
step: 420, loss: 0.00014248635852709413
step: 430, loss: 0.01292387954890728
step: 440, loss: 0.0007084644166752696
step: 450, loss: 0.0169801265001297
step: 460, loss: 0.0022845373023301363
step: 470, loss: 0.03307066485285759
step: 480, loss: 0.0017162057338282466
step: 490, loss: 0.0004711878136731684
step: 500, loss: 0.00758043909445405
step: 510, loss: 7.764818292343989e-05
step: 520, loss: 0.003437277628108859
step: 530, loss: 0.00014490554167423397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9363261566651397, f1=0.929539295392954,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010803166514961049
step: 10, loss: 0.0007698593544773757
step: 20, loss: 0.00010574371844995767
step: 30, loss: 0.0002629506925586611
step: 40, loss: 0.0029080528765916824
step: 50, loss: 0.006664243992418051
step: 60, loss: 0.00023969054745975882
step: 70, loss: 9.761020191945136e-05
step: 80, loss: 0.0009907055646181107
step: 90, loss: 6.194108573254198e-05
step: 100, loss: 0.0005231465911492705
step: 110, loss: 5.445455462904647e-05
step: 120, loss: 5.680859612766653e-05
step: 130, loss: 0.0005679209716618061
step: 140, loss: 0.006311539560556412
step: 150, loss: 0.0001107040952774696
step: 160, loss: 0.001405289862304926
step: 170, loss: 0.00014307748642750084
step: 180, loss: 8.0681755207479e-05
step: 190, loss: 0.0001203968349727802
step: 200, loss: 0.0019893525168299675
step: 210, loss: 7.155106140999123e-05
step: 220, loss: 0.00017683500482235104
step: 230, loss: 0.0008679481106810272

```



```

step: 240, loss: 0.03974875435233116
step: 250, loss: 2.530117126298137e-05
step: 260, loss: 4.2581999878166243e-05
step: 270, loss: 0.0034789356868714094
step: 280, loss: 2.822560963977594e-05
step: 290, loss: 8.733568392926827e-05
step: 300, loss: 0.00012165847147116438
step: 310, loss: 4.5502289140131325e-05
step: 320, loss: 0.0003845925675705075
step: 330, loss: 0.00032488928991369903
step: 340, loss: 2.8560780265252106e-05
step: 350, loss: 0.0004308696370571852
step: 360, loss: 0.1790219098329544
step: 370, loss: 0.00022537217591889203
step: 380, loss: 2.5342320441268384e-05
step: 390, loss: 0.0007381258183158934
step: 400, loss: 3.4165452234447e-05
step: 410, loss: 0.0009094882407225668
step: 420, loss: 8.583794260630384e-05
step: 430, loss: 0.0015593903372064233
step: 440, loss: 1.9028493625228293e-05
step: 450, loss: 3.4694843634497374e-05
step: 460, loss: 0.0001635332591831684
step: 470, loss: 2.762581243587192e-05
step: 480, loss: 0.00011449505109339952
step: 490, loss: 0.0001253961236216128
step: 500, loss: 0.000523582159075886
step: 510, loss: 0.00040718901436775923
step: 520, loss: 2.9550510589615442e-05
step: 530, loss: 3.853056114166975e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9371847776249428, f1=0.9320036264732547,
best_f1=0.9315693430656934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018632016144692898
step: 10, loss: 0.0002137235424015671
step: 20, loss: 6.397609831765294e-05
step: 30, loss: 9.198963380185887e-05
step: 40, loss: 9.301609679823741e-05
step: 50, loss: 0.00036642648046836257
step: 60, loss: 4.203159915050492e-05
step: 70, loss: 0.0008167946944013238
step: 80, loss: 9.316339128417894e-05
step: 90, loss: 3.0981755116954446e-05

```

```
step: 100, loss: 2.1710551664000377e-05
step: 110, loss: 0.02178286761045456
step: 120, loss: 0.0012934625847265124
step: 130, loss: 0.00011134196392958984
step: 140, loss: 2.750966996245552e-05
step: 150, loss: 0.009721823036670685
step: 160, loss: 2.7863301511388272e-05
step: 170, loss: 0.1303568333387375
step: 180, loss: 6.502383621409535e-05
step: 190, loss: 0.003823275677859783
step: 200, loss: 0.0001346818607999012
step: 210, loss: 3.063937765546143e-05
step: 220, loss: 8.825246914057061e-05
step: 230, loss: 0.0001370312529616058
step: 240, loss: 0.002894796896725893
step: 250, loss: 0.0004406192456372082
step: 260, loss: 0.00356482551433146
step: 270, loss: 0.04232731834053993
step: 280, loss: 0.001293362583965063
step: 290, loss: 0.001121865352615714
step: 300, loss: 0.012399734929203987
step: 310, loss: 0.0007858793251216412
step: 320, loss: 8.502578566549346e-05
step: 330, loss: 0.0003251872258260846
step: 340, loss: 0.000668641529045999
step: 350, loss: 0.00024149875389412045
step: 360, loss: 0.0002538806584198028
step: 370, loss: 9.315620263805613e-05
step: 380, loss: 4.972489841748029e-05
step: 390, loss: 8.028154115891084e-05
step: 400, loss: 0.00029633560916408896
step: 410, loss: 0.00028568715788424015
step: 420, loss: 0.00011800340143963695
step: 430, loss: 7.12380715413019e-05
step: 440, loss: 0.002133310539647937
step: 450, loss: 0.001188387512229383
step: 460, loss: 0.00026228761998936534
step: 470, loss: 0.00010197937808698043
step: 480, loss: 0.0002852431498467922
step: 490, loss: 0.0005354899913072586
step: 500, loss: 6.353646313073114e-05
step: 510, loss: 4.453347355592996e-05
step: 520, loss: 0.0011619824217632413
step: 530, loss: 8.469611202599481e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9371534195933457, f1=0.9333941605839416,
```

```

best_f1=0.9315693430656934
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:23, 240.10it/s]
load_f1 = 0.9368029739776952
real_f1 = 0.9362292051756008
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:18, 243.80it/s]

```

#### 4.1.4 Amazon-Google - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \

```

```

--task Structured/Amazon-Google \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True

```

```

master_weights      : True
loss_scale           : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8374397158622742
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06340006738901138
step: 20, loss: 0.3843121826648712
step: 30, loss: 0.3751490116119385
step: 40, loss: 0.5047157406806946
step: 50, loss: 0.3017215430736542
step: 60, loss: 0.36106443405151367
step: 70, loss: 0.25520798563957214
step: 80, loss: 0.286540150642395
step: 90, loss: 0.42371031641960144
step: 100, loss: 0.20895536243915558
step: 110, loss: 0.2971155345439911
step: 120, loss: 0.2219552993774414
step: 130, loss: 0.21786783635616302
step: 140, loss: 0.24419184029102325
step: 150, loss: 0.30384477972984314
step: 160, loss: 0.29298824071884155
step: 170, loss: 0.19324317574501038
step: 180, loss: 0.21304228901863098
step: 190, loss: 0.2200564593076706
step: 200, loss: 0.15153886377811432
step: 210, loss: 0.44598114490509033
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```

epoch 1: dev_f1=0.5900990099009902, f1=0.5708333333333334,
best_f1=0.5708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09750660508871078
step: 10, loss: 0.09025165438652039
step: 20, loss: 0.2729460895061493
step: 30, loss: 0.10083398967981339
step: 40, loss: 0.15119975805282593
step: 50, loss: 0.21151088178157806
step: 60, loss: 0.07216775417327881
step: 70, loss: 0.11337800323963165
step: 80, loss: 0.1788257658481598
step: 90, loss: 0.08721749484539032
step: 100, loss: 0.14566992223262787
step: 110, loss: 0.09309547394514084
step: 120, loss: 0.17573915421962738
step: 130, loss: 0.2128618061542511
step: 140, loss: 0.20536789298057556
step: 150, loss: 0.2655814588069916
step: 160, loss: 0.21146981418132782
step: 170, loss: 0.3067380487918854
step: 180, loss: 0.22952638566493988
step: 190, loss: 0.1790381222963333
step: 200, loss: 0.2267795354127884
step: 210, loss: 0.21082091331481934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6741996233521658, f1=0.6233269598470363,
best_f1=0.6233269598470363
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1323588639497757
step: 10, loss: 0.19872474670410156
step: 20, loss: 0.30721569061279297
step: 30, loss: 0.10010521858930588
step: 40, loss: 0.18390776216983795
step: 50, loss: 0.10981401056051254
step: 60, loss: 0.2644250690937042
step: 70, loss: 0.15004661679267883
step: 80, loss: 0.09615789353847504

```

```

step: 90, loss: 0.13076207041740417
step: 100, loss: 0.03359689936041832
step: 110, loss: 0.16432258486747742
step: 120, loss: 0.22390665113925934
step: 130, loss: 0.14752808213233948
step: 140, loss: 0.2049427628517151
step: 150, loss: 0.15694233775138855
step: 160, loss: 0.1174248456954956
step: 170, loss: 0.1683119535446167
step: 180, loss: 0.10062920302152634
step: 190, loss: 0.165788933634758
step: 200, loss: 0.1317332535982132
step: 210, loss: 0.17583811283111572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6818181818181818, f1=0.6565464895635673,
best_f1=0.6565464895635673
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14487212896347046
step: 10, loss: 0.013683976605534554
step: 20, loss: 0.12877073884010315
step: 30, loss: 0.07900840789079666
step: 40, loss: 0.07075915485620499
step: 50, loss: 0.05022590979933739
step: 60, loss: 0.0484403558075428
step: 70, loss: 0.193324476480484
step: 80, loss: 0.2679404616355896
step: 90, loss: 0.03485006093978882
step: 100, loss: 0.10446594655513763
step: 110, loss: 0.08528883755207062
step: 120, loss: 0.0190426092594862
step: 130, loss: 0.08886898308992386
step: 140, loss: 0.193056121468544
step: 150, loss: 0.3879421651363373
step: 160, loss: 0.07186275720596313
step: 170, loss: 0.1260625422000885
step: 180, loss: 0.1557256430387497
step: 190, loss: 0.13580067455768585
step: 200, loss: 0.10406650602817535
step: 210, loss: 0.030354365706443787
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.673866090712743, f1=0.6129753914988815,
best_f1=0.6565464895635673
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08820614963769913
step: 10, loss: 0.1597777158021927
step: 20, loss: 0.0911073386669159
step: 30, loss: 0.04906332120299339
step: 40, loss: 0.1703072041273117
step: 50, loss: 0.1721193641424179
step: 60, loss: 0.023857133463025093
step: 70, loss: 0.2566497027873993
step: 80, loss: 0.048662081360816956
step: 90, loss: 0.05017467215657234
step: 100, loss: 0.06974989920854568
step: 110, loss: 0.017872627824544907
step: 120, loss: 0.010817239992320538
step: 130, loss: 0.04583116993308067
step: 140, loss: 0.12706682085990906
step: 150, loss: 0.02310236543416977
step: 160, loss: 0.041521307080984116
step: 170, loss: 0.07308269292116165
step: 180, loss: 0.13501086831092834
step: 190, loss: 0.11541935056447983
step: 200, loss: 0.2111191749572754
step: 210, loss: 0.2483721226453781
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.6827458256029686, f1=0.650557620817844,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07819691300392151
step: 10, loss: 0.03852103278040886
step: 20, loss: 0.03637200966477394
step: 30, loss: 0.1707315295934677
step: 40, loss: 0.07786399871110916
step: 50, loss: 0.07465296238660812
step: 60, loss: 0.08833255618810654

```



```

step: 70, loss: 0.030368562787771225
step: 80, loss: 0.1008290946483612
step: 90, loss: 0.0655386745929718
step: 100, loss: 0.08040983229875565
step: 110, loss: 0.008026291616261005
step: 120, loss: 0.09300704300403595
step: 130, loss: 0.011188705451786518
step: 140, loss: 0.08848079293966293
step: 150, loss: 0.05071326345205307
step: 160, loss: 0.17308409512043
step: 170, loss: 0.07570935040712357
step: 180, loss: 0.009092312306165695
step: 190, loss: 0.12425000220537186
step: 200, loss: 0.013257211074233055
step: 210, loss: 0.028415221720933914
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6444007858546168, f1=0.6382978723404256,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030182527378201485
step: 10, loss: 0.01875421218574047
step: 20, loss: 0.03442048281431198
step: 30, loss: 0.013110090978443623
step: 40, loss: 0.07701827585697174
step: 50, loss: 0.20681017637252808
step: 60, loss: 0.2069692313671112
step: 70, loss: 0.04220178350806236
step: 80, loss: 0.07386112958192825
step: 90, loss: 0.005200638435781002
step: 100, loss: 0.04863280802965164
step: 110, loss: 0.133165642619133
step: 120, loss: 0.3602154850959778
step: 130, loss: 0.026968419551849365
step: 140, loss: 0.0034950615372508764
step: 150, loss: 0.01320751290768385
step: 160, loss: 0.015146108344197273
step: 170, loss: 0.02710852026939392
step: 180, loss: 0.00872081145644188
step: 190, loss: 0.1330300271511078
step: 200, loss: 0.0684085339307785
step: 210, loss: 0.11657299846410751
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 7: dev_f1=0.6465028355387523, f1=0.6285714285714287,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.051129553467035294
step: 10, loss: 0.013569018803536892
step: 20, loss: 0.06441193073987961
step: 30, loss: 0.16761261224746704
step: 40, loss: 0.059572335332632065
step: 50, loss: 0.06876415759325027
step: 60, loss: 0.18685071170330048
step: 70, loss: 0.011200048960745335
step: 80, loss: 0.057330455631017685
step: 90, loss: 0.018037280067801476
step: 100, loss: 0.08701155334711075
step: 110, loss: 0.00783547293394804
step: 120, loss: 0.032038796693086624
step: 130, loss: 0.09495491534471512
step: 140, loss: 0.006196354981511831
step: 150, loss: 0.02238890342414379
step: 160, loss: 0.03992573916912079
step: 170, loss: 0.015163905918598175
step: 180, loss: 0.1332373023033142
step: 190, loss: 0.055458858609199524
step: 200, loss: 0.15920545160770416
step: 210, loss: 0.22046269476413727
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.6551059730250481, f1=0.6410748560460652,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007373481523245573
step: 10, loss: 0.08606186509132385
step: 20, loss: 0.11273185163736343
step: 30, loss: 0.20175492763519287
step: 40, loss: 0.007373409811407328
step: 50, loss: 0.0026969104073941708
step: 60, loss: 0.0033731781877577305
step: 70, loss: 0.0316268652677536
step: 80, loss: 0.01986994594335556
step: 90, loss: 0.05770502984523773
step: 100, loss: 0.01368036586791277
step: 110, loss: 0.0026456378400325775
step: 120, loss: 0.004598108120262623

```

```

step: 130, loss: 0.0031794854439795017
step: 140, loss: 0.01840228959918022
step: 150, loss: 0.07914955168962479
step: 160, loss: 0.0016194795025512576
step: 170, loss: 0.020273737609386444
step: 180, loss: 0.06134679168462753
step: 190, loss: 0.2076578140258789
step: 200, loss: 0.025403469800949097
step: 210, loss: 0.04343825578689575
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6512524084778419, f1=0.6242544731610339,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02747523784637451
step: 10, loss: 0.012573963031172752
step: 20, loss: 0.0030664089135825634
step: 30, loss: 0.06993118673563004
step: 40, loss: 0.16731829941272736
step: 50, loss: 0.0013975307811051607
step: 60, loss: 0.005251326598227024
step: 70, loss: 0.08989956229925156
step: 80, loss: 0.007178361993283033
step: 90, loss: 0.024442389607429504
step: 100, loss: 0.004291769582778215
step: 110, loss: 0.019226741045713425
step: 120, loss: 0.006113544572144747
step: 130, loss: 0.021615395322442055
step: 140, loss: 0.0012543441262096167
step: 150, loss: 0.02390589751303196
step: 160, loss: 0.07521573454141617
step: 170, loss: 0.02840963937342167
step: 180, loss: 0.016246305778622627
step: 190, loss: 0.14540398120880127
step: 200, loss: 0.04302439093589783
step: 210, loss: 0.012646103277802467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6449136276391555, f1=0.6501901140684411,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011520742438733578

```

```

step: 10, loss: 0.07128393650054932
step: 20, loss: 0.00592435710132122
step: 30, loss: 0.05106835439801216
step: 40, loss: 0.040937792509794235
step: 50, loss: 0.02435430698096752
step: 60, loss: 0.044570378959178925
step: 70, loss: 0.0358433723449707
step: 80, loss: 0.05048811435699463
step: 90, loss: 0.01570173166692257
step: 100, loss: 0.010175746865570545
step: 110, loss: 0.030998682603240013
step: 120, loss: 0.03807896375656128
step: 130, loss: 0.05728282034397125
step: 140, loss: 0.22140656411647797
step: 150, loss: 0.030800579115748405
step: 160, loss: 0.013789542019367218
step: 170, loss: 0.10577879846096039
step: 180, loss: 0.08432025462388992
step: 190, loss: 0.01666196621954441
step: 200, loss: 0.0005850215093232691
step: 210, loss: 0.006059026811271906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6313645621181263, f1=0.6414342629482072,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004120181780308485
step: 10, loss: 0.005673200823366642
step: 20, loss: 0.026526199653744698
step: 30, loss: 0.0035969377495348454
step: 40, loss: 0.13170349597930908
step: 50, loss: 0.019932448863983154
step: 60, loss: 0.10575709491968155
step: 70, loss: 0.01975102163851261
step: 80, loss: 0.058103807270526886
step: 90, loss: 0.003645826829597354
step: 100, loss: 0.0019832102116197348
step: 110, loss: 0.002906502690166235
step: 120, loss: 0.0030734131578356028
step: 130, loss: 0.013986286707222462
step: 140, loss: 0.16472186148166656
step: 150, loss: 0.001957800704985857
step: 160, loss: 0.026877857744693756
step: 170, loss: 0.0018977316794916987
step: 180, loss: 0.052454520016908646

```

```

step: 190, loss: 0.08916796743869781
step: 200, loss: 0.02382560260593891
step: 210, loss: 0.04866573214530945
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6247818499127399, f1=0.618374558303887,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00380169739946723
step: 10, loss: 0.0036795292980968952
step: 20, loss: 0.00156933325342834
step: 30, loss: 0.031694650650024414
step: 40, loss: 0.0007970555452629924
step: 50, loss: 0.00847212690860033
step: 60, loss: 0.003024404402822256
step: 70, loss: 0.08786965161561966
step: 80, loss: 0.005358377005904913
step: 90, loss: 0.027824660763144493
step: 100, loss: 0.024345654994249344
step: 110, loss: 0.05762968212366104
step: 120, loss: 0.004230013117194176
step: 130, loss: 0.0016498122131451964
step: 140, loss: 0.018440505489706993
step: 150, loss: 0.00147564010694623
step: 160, loss: 0.03435566648840904
step: 170, loss: 0.0010688253678381443
step: 180, loss: 0.13034294545650482
step: 190, loss: 0.004731033928692341
step: 200, loss: 0.0022355024702847004
step: 210, loss: 0.05024878680706024
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6403326403326404, f1=0.6285714285714287,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022137990221381187
step: 10, loss: 0.008586747571825981
step: 20, loss: 0.001336292247287929
step: 30, loss: 0.004211194347590208
step: 40, loss: 0.008288680575788021
step: 50, loss: 0.04505493491888046
step: 60, loss: 0.006617838982492685

```

```

step: 70, loss: 0.014860352501273155
step: 80, loss: 0.036814186722040176
step: 90, loss: 0.1045897901058197
step: 100, loss: 0.0011181135196238756
step: 110, loss: 0.02344008907675743
step: 120, loss: 0.003891418222337961
step: 130, loss: 0.022531934082508087
step: 140, loss: 0.0016343464376404881
step: 150, loss: 0.011472453363239765
step: 160, loss: 0.11238051950931549
step: 170, loss: 0.08132405579090118
step: 180, loss: 0.02985074184834957
step: 190, loss: 0.0058690402656793594
step: 200, loss: 0.11225888878107071
step: 210, loss: 0.05913335457444191
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6320000000000001, f1=0.6247544204322201,
best_f1=0.650557620817844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008288910612463951
step: 10, loss: 0.018128033727407455
step: 20, loss: 0.0042546652257442474
step: 30, loss: 0.016366705298423767
step: 40, loss: 0.0026210444048047066
step: 50, loss: 0.13899409770965576
step: 60, loss: 0.006856098771095276
step: 70, loss: 0.0029259666334837675
step: 80, loss: 0.07524749636650085
step: 90, loss: 0.008776630274951458
step: 100, loss: 0.02466822788119316
step: 110, loss: 0.0011000708909705281
step: 120, loss: 0.003300655400380492
step: 130, loss: 0.2937827706336975
step: 140, loss: 0.01658564805984497
step: 150, loss: 0.039009254425764084
step: 160, loss: 0.03648149222135544
step: 170, loss: 0.08467308431863785
step: 180, loss: 0.001894832239486277
step: 190, loss: 0.020449966192245483
step: 200, loss: 0.018083691596984863
step: 210, loss: 0.04150353744626045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```
epoch 15: dev_f1=0.6317907444668007, f1=0.6336633663366337,
best_f1=0.650557620817844
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2: FP16 training with FP32 batchnorm and FP32
master weights.
```

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
2293it [00:06, 328.37it/s]
load_f1 = 0.6679035250463822
real_f1 = 0.6654205607476634
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
4398it [00:17, 249.01it/s]
```

#### 4.1.5 Walmart-Amazon - Running the matcher

```
[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
```

```

!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Walmart-Amazon \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/Walmart-Amazon \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False

```



```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8521410822868347
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.17649491131305695
step: 20, loss: 0.15465934574604034
step: 30, loss: 0.5011898279190063
step: 40, loss: 0.25778332352638245
step: 50, loss: 0.3056910037994385
step: 60, loss: 0.36232903599739075
step: 70, loss: 0.1830035001039505
step: 80, loss: 0.5377705693244934
step: 90, loss: 0.24664849042892456
step: 100, loss: 0.2372722029685974
step: 110, loss: 0.2369520217180252
step: 120, loss: 0.40680915117263794
step: 130, loss: 0.3456183075904846
step: 140, loss: 0.3250541687011719
step: 150, loss: 0.26447227597236633
step: 160, loss: 0.21754983067512512
step: 170, loss: 0.37251079082489014
step: 180, loss: 0.2811882197856903
step: 190, loss: 0.139614999294281
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.5822102425876011, f1=0.5872576177285318,

```

```

best_f1=0.5872576177285318
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25033852458000183
step: 10, loss: 0.04671883210539818
step: 20, loss: 0.0354493223130703
step: 30, loss: 0.12505513429641724
step: 40, loss: 0.3925721347332001
step: 50, loss: 0.2825397551059723
step: 60, loss: 0.19084244966506958
step: 70, loss: 0.21804478764533997
step: 80, loss: 0.155877485871315
step: 90, loss: 0.135179340839386
step: 100, loss: 0.2874252200126648
step: 110, loss: 0.05210914462804794
step: 120, loss: 0.18566858768463135
step: 130, loss: 0.08482477068901062
step: 140, loss: 0.0923910066485405
step: 150, loss: 0.030816856771707535
step: 160, loss: 0.12051979452371597
step: 170, loss: 0.13736896216869354
step: 180, loss: 0.0709504783153534
step: 190, loss: 0.1475544273853302
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7648725212464589, f1=0.7688022284122563,
best_f1=0.7688022284122563
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06093107908964157
step: 10, loss: 0.21226245164871216
step: 20, loss: 0.056839074939489365
step: 30, loss: 0.07728437334299088
step: 40, loss: 0.04570504277944565
step: 50, loss: 0.24158193171024323
step: 60, loss: 0.04484468325972557
step: 70, loss: 0.10532929748296738
step: 80, loss: 0.2689704895019531
step: 90, loss: 0.092542864382267
step: 100, loss: 0.1290816217660904
step: 110, loss: 0.2362891584634781

```

```

step: 120, loss: 0.008756259456276894
step: 130, loss: 0.018028615042567253
step: 140, loss: 0.12111889570951462
step: 150, loss: 0.0630468800663948
step: 160, loss: 0.13976743817329407
step: 170, loss: 0.01976645737886429
step: 180, loss: 0.05753309279680252
step: 190, loss: 0.06612929701805115
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7727272727272727, f1=0.744047619047619,
best_f1=0.744047619047619
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12015779316425323
step: 10, loss: 0.03232123330235481
step: 20, loss: 0.010923270136117935
step: 30, loss: 0.018419530242681503
step: 40, loss: 0.0036711355205625296
step: 50, loss: 0.010118352249264717
step: 60, loss: 0.17565998435020447
step: 70, loss: 0.00608109962195158
step: 80, loss: 0.21780458092689514
step: 90, loss: 0.01360237691551447
step: 100, loss: 0.025743378326296806
step: 110, loss: 0.03438364341855049
step: 120, loss: 0.13248765468597412
step: 130, loss: 0.14957384765148163
step: 140, loss: 0.08194922655820847
step: 150, loss: 0.046204667538404465
step: 160, loss: 0.04426560178399086
step: 170, loss: 0.013750974088907242
step: 180, loss: 0.0653756707906723
step: 190, loss: 0.14492182433605194
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8067226890756303, f1=0.7910863509749304,

```

```

best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05159483477473259
step: 10, loss: 0.007397535257041454
step: 20, loss: 0.03811151534318924
step: 30, loss: 0.0025839584413915873
step: 40, loss: 0.05122529715299606
step: 50, loss: 0.008331668563187122
step: 60, loss: 0.016420481726527214
step: 70, loss: 0.016766207292675972
step: 80, loss: 0.020309561863541603
step: 90, loss: 0.016227208077907562
step: 100, loss: 0.09375530481338501
step: 110, loss: 0.011685702949762344
step: 120, loss: 0.0008269163081422448
step: 130, loss: 0.004261649213731289
step: 140, loss: 0.005343222059309483
step: 150, loss: 0.02957608923316002
step: 160, loss: 0.0060256957076489925
step: 170, loss: 0.06515932828187943
step: 180, loss: 0.20773610472679138
step: 190, loss: 0.10620531439781189
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7859078590785908, f1=0.7650273224043715,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2166372388601303
step: 10, loss: 0.016053972765803337
step: 20, loss: 0.0033012356143444777
step: 30, loss: 0.061037324368953705
step: 40, loss: 0.02393646165728569
step: 50, loss: 0.003989117685705423
step: 60, loss: 0.0034869385417550802
step: 70, loss: 0.028741255402565002
step: 80, loss: 0.04067056626081467
step: 90, loss: 0.15991365909576416
step: 100, loss: 0.05122220888733864
step: 110, loss: 0.048713140189647675
step: 120, loss: 0.10892705619335175
step: 130, loss: 0.0059663294814527035
step: 140, loss: 0.005935473833233118
step: 150, loss: 0.05490309000015259

```

```

step: 160, loss: 0.015057059936225414
step: 170, loss: 0.03103581629693508
step: 180, loss: 0.015948403626680374
step: 190, loss: 0.018916623666882515
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7754010695187165, f1=0.7769028871391076,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019112879410386086
step: 10, loss: 0.06379923224449158
step: 20, loss: 0.006282787304371595
step: 30, loss: 0.01033988781273365
step: 40, loss: 0.007815835997462273
step: 50, loss: 0.04400308057665825
step: 60, loss: 0.030747152864933014
step: 70, loss: 0.0023807790130376816
step: 80, loss: 0.015855230391025543
step: 90, loss: 0.003129858523607254
step: 100, loss: 0.005243821069598198
step: 110, loss: 0.003578626550734043
step: 120, loss: 0.010540034621953964
step: 130, loss: 0.001628225203603506
step: 140, loss: 0.0019396692514419556
step: 150, loss: 0.02283506840467453
step: 160, loss: 0.000823179492726922
step: 170, loss: 0.0037410056684166193
step: 180, loss: 0.02563267946243286
step: 190, loss: 0.03320802003145218
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7875354107648725, f1=0.7675070028011204,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006727385334670544
step: 10, loss: 0.018853342160582542
step: 20, loss: 0.002267244504764676
step: 30, loss: 0.00613250071182847
step: 40, loss: 0.04959089308977127
step: 50, loss: 0.0027470928616821766
step: 60, loss: 0.00701146712526679
step: 70, loss: 0.006922421511262655

```

```

step: 80, loss: 0.11088845133781433
step: 90, loss: 0.001745451707392931
step: 100, loss: 0.08182837069034576
step: 110, loss: 0.004394145216792822
step: 120, loss: 0.05019668489694595
step: 130, loss: 0.0036748440470546484
step: 140, loss: 0.004348506685346365
step: 150, loss: 0.10782849788665771
step: 160, loss: 0.005187273025512695
step: 170, loss: 0.0005894224159419537
step: 180, loss: 0.0221764724701643
step: 190, loss: 0.07744207233190536
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7999999999999999, f1=0.7893333333333333,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001296707196161151
step: 10, loss: 0.085373155772686
step: 20, loss: 0.026808004826307297
step: 30, loss: 0.00422366289421916
step: 40, loss: 0.01102062314748764
step: 50, loss: 0.0021479029674082994
step: 60, loss: 0.0009299708181060851
step: 70, loss: 0.0018586921505630016
step: 80, loss: 0.003549773246049881
step: 90, loss: 0.01126942504197359
step: 100, loss: 0.005217397585511208
step: 110, loss: 0.020932350307703018
step: 120, loss: 0.02126036398112774
step: 130, loss: 0.0016441107727587223
step: 140, loss: 0.0009666977566666901
step: 150, loss: 0.11784065514802933
step: 160, loss: 0.09826062619686127
step: 170, loss: 0.006982899736613035
step: 180, loss: 0.006759869400411844
step: 190, loss: 0.004220960661768913
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7833333333333333, f1=0.7611111111111111,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.0002808751305565238
step: 10, loss: 0.0765058621764183
step: 20, loss: 0.0020954706706106663
step: 30, loss: 0.004203233402222395
step: 40, loss: 0.0041753011755645275
step: 50, loss: 0.0012972961412742734
step: 60, loss: 0.015482166782021523
step: 70, loss: 0.01267923042178154
step: 80, loss: 0.0043167113326489925
step: 90, loss: 0.02029411308467388
step: 100, loss: 0.011191776022315025
step: 110, loss: 0.016141420230269432
step: 120, loss: 0.016630303114652634
step: 130, loss: 0.18118208646774292
step: 140, loss: 0.003846118226647377
step: 150, loss: 0.0008113498915918171
step: 160, loss: 0.0008298925822600722
step: 170, loss: 0.008420969359576702
step: 180, loss: 0.018973320722579956
step: 190, loss: 0.0010679980041459203
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.8010899182561309, f1=0.7989276139410189,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00252813589759171
step: 10, loss: 0.0005024529527872801
step: 20, loss: 0.06522252410650253
step: 30, loss: 0.002492654835805297
step: 40, loss: 0.01601376384496689
step: 50, loss: 0.011126954108476639
step: 60, loss: 0.00037768183392472565
step: 70, loss: 0.003670288482680917
step: 80, loss: 0.0011846267152577639
step: 90, loss: 0.00197302782908082
step: 100, loss: 0.0007007595268078148
step: 110, loss: 0.0014175019459798932
step: 120, loss: 0.009903619065880775
step: 130, loss: 0.0006632799049839377
step: 140, loss: 0.0010997053468599916
step: 150, loss: 0.00034604151733219624
step: 160, loss: 0.0007403666386380792
step: 170, loss: 0.0013953212182968855
step: 180, loss: 0.00959487073123455
step: 190, loss: 0.007733520120382309

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7821229050279329, f1=0.776536312849162,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006582410074770451
step: 10, loss: 0.016820233315229416
step: 20, loss: 0.0027161596808582544
step: 30, loss: 0.0023971020709723234
step: 40, loss: 0.012341088615357876
step: 50, loss: 0.0008720112964510918
step: 60, loss: 0.0012001779396086931
step: 70, loss: 0.00687860744073987
step: 80, loss: 0.0006486852653324604
step: 90, loss: 0.0015830707270652056
step: 100, loss: 0.000812419515568763
step: 110, loss: 0.001241039833985269
step: 120, loss: 0.0009134691790677607
step: 130, loss: 0.03117840550839901
step: 140, loss: 0.0009679551003500819
step: 150, loss: 0.0009596521849744022
step: 160, loss: 0.0013778172433376312
step: 170, loss: 0.00944607425481081
step: 180, loss: 0.0007370450184680521
step: 190, loss: 0.0030625630170106888
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7870619946091645, f1=0.7893333333333333,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005046098958700895
step: 10, loss: 0.011088780127465725
step: 20, loss: 0.00024952678359113634
step: 30, loss: 0.04461977630853653
step: 40, loss: 0.0021872390061616898
step: 50, loss: 0.00044688963680528104
step: 60, loss: 0.0007300105644389987
step: 70, loss: 0.046793825924396515
step: 80, loss: 0.001402552123181522
step: 90, loss: 0.0007357667200267315
step: 100, loss: 0.000588340568356216
step: 110, loss: 0.0020939207170158625

```



```

step: 120, loss: 0.002980967052280903
step: 130, loss: 0.0057360269129276276
step: 140, loss: 0.0009842092404142022
step: 150, loss: 0.00453893793746829
step: 160, loss: 0.00041972106555476785
step: 170, loss: 0.001349522382952273
step: 180, loss: 0.01971304789185524
step: 190, loss: 0.0005938482936471701
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7874015748031497, f1=0.7886597938144331,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001209631678648293
step: 10, loss: 0.0008025916176848114
step: 20, loss: 0.00675413990393281
step: 30, loss: 0.007489008363336325
step: 40, loss: 0.0006845346069894731
step: 50, loss: 0.027845468372106552
step: 60, loss: 0.010467217303812504
step: 70, loss: 0.00446954183280468
step: 80, loss: 0.0005436034989543259
step: 90, loss: 0.0018154458375647664
step: 100, loss: 0.0007754412945359945
step: 110, loss: 0.00028402978205122054
step: 120, loss: 0.015403049997985363
step: 130, loss: 0.0022760762367397547
step: 140, loss: 0.0015172016574069858
step: 150, loss: 0.006774043198674917
step: 160, loss: 0.0025874560233205557
step: 170, loss: 0.00033210183028131723
step: 180, loss: 0.00915850605815649
step: 190, loss: 0.1564178466796875
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7843137254901961, f1=0.7740112994350282,
best_f1=0.7910863509749304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012643240625038743
step: 10, loss: 0.0016796853160485625
step: 20, loss: 0.10487577319145203
step: 30, loss: 0.006149631924927235

```

```

step: 40, loss: 0.000473206106107682
step: 50, loss: 0.0015443396987393498
step: 60, loss: 0.000614830874837935
step: 70, loss: 0.001689039752818644
step: 80, loss: 0.0012237413320690393
step: 90, loss: 0.05322568118572235
step: 100, loss: 0.0028824685141444206
step: 110, loss: 0.0005522092105820775
step: 120, loss: 0.027011973783373833
step: 130, loss: 0.002239271765574813
step: 140, loss: 0.019057391211390495
step: 150, loss: 0.0008721628692001104
step: 160, loss: 0.005444167647510767
step: 170, loss: 0.0009241392253898084
step: 180, loss: 0.0015101950848475099
step: 190, loss: 0.03375427797436714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7821229050279329, f1=0.7787114845938375,
best_f1=0.7910863509749304
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2049it [00:09, 221.30it/s]
load_f1 = 0.5877551020408164
real_f1 = 0.5777777777777778
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:18, 236.75it/s]

```

## 4.2 DITTO TEXTUAL

### 4.2.1 Abt-Buy - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Textual/Abt-Buy \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize \
  --da entry_swap \
  --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Textual/Abt-Buy \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \

```

```
--fp16 \  
--checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True  
opt_level        : 02  
cast_model_type  : torch.float16  
patch_torch_functions : False  
keep_batchnorm_fp32 : True  
master_weights   : True  
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True  
opt_level        : 02  
cast_model_type  : torch.float16  
patch_torch_functions : False  
keep_batchnorm_fp32 : True  
master_weights   : True  
loss_scale       : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")  
/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0  
/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",  
UserWarning)
```

```
step: 0, loss: 0.8373283743858337
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.22078442573547363
```

```
step: 20, loss: 0.1550557017326355
```

```
step: 30, loss: 0.22819335758686066
```

```
step: 40, loss: 0.3157814145088196
```

```
step: 50, loss: 0.3747500479221344
```

```
step: 60, loss: 0.44327181577682495
```

```

step: 70, loss: 0.3132014274597168
step: 80, loss: 0.2501460611820221
step: 90, loss: 0.41227269172668457
step: 100, loss: 0.23804733157157898
step: 110, loss: 0.18422962725162506
step: 120, loss: 0.5499745011329651
step: 130, loss: 0.4175878167152405
step: 140, loss: 0.470758318901062
step: 150, loss: 0.12094946950674057
step: 160, loss: 0.29890406131744385
step: 170, loss: 0.15686346590518951
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6470588235294118, f1=0.6755555555555557,
best_f1=0.6755555555555557
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2386687695980072
step: 10, loss: 0.08667623996734619
step: 20, loss: 0.25153353810310364
step: 30, loss: 0.21462984383106232
step: 40, loss: 0.11965291947126389
step: 50, loss: 0.25512391328811646
step: 60, loss: 0.1556556671857834
step: 70, loss: 0.3494889736175537
step: 80, loss: 0.1588650345802307
step: 90, loss: 0.23275236785411835
step: 100, loss: 0.1808776706457138
step: 110, loss: 0.14008383452892303
step: 120, loss: 0.04706773906946182
step: 130, loss: 0.11270874738693237
step: 140, loss: 0.16224506497383118
step: 150, loss: 0.24508483707904816
step: 160, loss: 0.21467365324497223
step: 170, loss: 0.24272848665714264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```

epoch 2: dev_f1=0.7394209354120268, f1=0.7477876106194691,
best_f1=0.7477876106194691
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07589399814605713
step: 10, loss: 0.1318753957748413
step: 20, loss: 0.10483599454164505
step: 30, loss: 0.17690584063529968
step: 40, loss: 0.010761265642940998
step: 50, loss: 0.17409873008728027
step: 60, loss: 0.0875474289059639
step: 70, loss: 0.07030701637268066
step: 80, loss: 0.22911730408668518
step: 90, loss: 0.07506527751684189
step: 100, loss: 0.02961614914238453
step: 110, loss: 0.03130844980478287
step: 120, loss: 0.006952475756406784
step: 130, loss: 0.1580847203731537
step: 140, loss: 0.014039678499102592
step: 150, loss: 0.07412687689065933
step: 160, loss: 0.03452194482088089
step: 170, loss: 0.18177668750286102
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7681159420289854, f1=0.7638888888888889,
best_f1=0.7638888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11864852905273438
step: 10, loss: 0.2629675567150116
step: 20, loss: 0.07542330771684647
step: 30, loss: 0.08982741087675095
step: 40, loss: 0.02523079514503479
step: 50, loss: 0.08917049318552017
step: 60, loss: 0.11689695715904236
step: 70, loss: 0.046360645443201065
step: 80, loss: 0.062077950686216354
step: 90, loss: 0.03747265413403511
step: 100, loss: 0.013376439921557903
step: 110, loss: 0.11754103004932404
step: 120, loss: 0.12758587300777435

```

```

step: 130, loss: 0.06397272646427155
step: 140, loss: 0.04023505002260208
step: 150, loss: 0.05860121175646782
step: 160, loss: 0.08333516120910645
step: 170, loss: 0.0138015141710639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7788461538461539, f1=0.772093023255814,
best_f1=0.772093023255814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17942054569721222
step: 10, loss: 0.05475874990224838
step: 20, loss: 0.009836561977863312
step: 30, loss: 0.10923176258802414
step: 40, loss: 0.0073710279539227486
step: 50, loss: 0.18622760474681854
step: 60, loss: 0.02417903020977974
step: 70, loss: 0.03867581859230995
step: 80, loss: 0.03873952478170395
step: 90, loss: 0.09017188102006912
step: 100, loss: 0.07184372842311859
step: 110, loss: 0.23231136798858643
step: 120, loss: 0.04647257924079895
step: 130, loss: 0.09069329500198364
step: 140, loss: 0.017671113833785057
step: 150, loss: 0.04966855049133301
step: 160, loss: 0.012660503387451172
step: 170, loss: 0.08673395216464996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7659574468085106, f1=0.7850467289719626,
best_f1=0.772093023255814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08046761155128479
step: 10, loss: 0.03758110851049423
step: 20, loss: 0.00907520018517971
step: 30, loss: 0.031937532126903534
step: 40, loss: 0.008146335370838642

```

```

step: 50, loss: 0.057075127959251404
step: 60, loss: 0.04692407324910164
step: 70, loss: 0.08600953966379166
step: 80, loss: 0.1787576675415039
step: 90, loss: 0.010542384348809719
step: 100, loss: 0.06736019998788834
step: 110, loss: 0.04431256279349327
step: 120, loss: 0.046864982694387436
step: 130, loss: 0.24688707292079926
step: 140, loss: 0.06268978863954544
step: 150, loss: 0.2512899339199066
step: 160, loss: 0.07104098051786423
step: 170, loss: 0.018147312104701996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7660550458715597, f1=0.7829977628635346,
best_f1=0.772093023255814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014868291094899178
step: 10, loss: 0.002198902890086174
step: 20, loss: 0.06205258145928383
step: 30, loss: 0.037835024297237396
step: 40, loss: 0.014371467754244804
step: 50, loss: 0.005175701342523098
step: 60, loss: 0.18864189088344574
step: 70, loss: 0.009144539013504982
step: 80, loss: 0.009496237151324749
step: 90, loss: 0.04051212966442108
step: 100, loss: 0.009729934856295586
step: 110, loss: 0.009417830966413021
step: 120, loss: 0.23907290399074554
step: 130, loss: 0.17894645035266876
step: 140, loss: 0.047512102872133255
step: 150, loss: 0.1380145251750946
step: 160, loss: 0.017099449411034584
step: 170, loss: 0.08524232357740402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.7902439024390243, f1=0.7795823665893272,
best_f1=0.7795823665893272

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010812409222126007
step: 10, loss: 0.008378490805625916
step: 20, loss: 0.020155327394604683
step: 30, loss: 0.071833036839962
step: 40, loss: 0.01083410158753395
step: 50, loss: 0.0030013183131814003
step: 60, loss: 0.010106462985277176
step: 70, loss: 0.07627581059932709
step: 80, loss: 0.01906459778547287
step: 90, loss: 0.09976047277450562
step: 100, loss: 0.030851731076836586
step: 110, loss: 0.03036927618086338
step: 120, loss: 0.003518862882629037
step: 130, loss: 0.0017128039617091417
step: 140, loss: 0.046701669692993164
step: 150, loss: 0.04909748584032059
step: 160, loss: 0.05626361072063446
step: 170, loss: 0.011165287345647812
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7821782178217821, f1=0.7801418439716311,
best_f1=0.7795823665893272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01189552154392004
step: 10, loss: 0.0030165244825184345
step: 20, loss: 0.005610199645161629
step: 30, loss: 0.011986706405878067
step: 40, loss: 0.14538757503032684
step: 50, loss: 0.08233707398176193
step: 60, loss: 0.015607321634888649
step: 70, loss: 0.0535256564617157
step: 80, loss: 0.11105892807245255
step: 90, loss: 0.09309419989585876
step: 100, loss: 0.03378951549530029
step: 110, loss: 0.013439501635730267
step: 120, loss: 0.005015576723963022
step: 130, loss: 0.0038802882190793753
step: 140, loss: 0.020046817138791084
step: 150, loss: 0.026896115392446518
step: 160, loss: 0.010719148442149162
step: 170, loss: 0.01583106815814972
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7893462469733656, f1=0.7808219178082191,
best_f1=0.7795823665893272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015405566664412618
step: 10, loss: 0.014745328575372696
step: 20, loss: 0.08655992150306702
step: 30, loss: 0.000957745942287147
step: 40, loss: 0.0950472503900528
step: 50, loss: 0.01598561741411686
step: 60, loss: 0.0058637382462620735
step: 70, loss: 0.26989084482192993
step: 80, loss: 0.10465515404939651
step: 90, loss: 0.047609198838472366
step: 100, loss: 0.06351587921380997
step: 110, loss: 0.0775756910443306
step: 120, loss: 0.03488229215145111
step: 130, loss: 0.021980859339237213
step: 140, loss: 0.0033784559927880764
step: 150, loss: 0.03250312805175781
step: 160, loss: 0.06530922651290894
step: 170, loss: 0.02231639437377453
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7880299251870324, f1=0.8018867924528301,
best_f1=0.7795823665893272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006302273832261562
step: 10, loss: 0.12738288938999176
step: 20, loss: 0.09026841819286346
step: 30, loss: 0.005226228851824999
step: 40, loss: 0.007654104381799698
step: 50, loss: 0.0013744377065449953
step: 60, loss: 0.04703064635396004
step: 70, loss: 0.0032123231794685125
step: 80, loss: 0.0017790694255381823
step: 90, loss: 0.002027529524639249
step: 100, loss: 0.004004445392638445
step: 110, loss: 0.010787680745124817
step: 120, loss: 0.019738899543881416
step: 130, loss: 0.0115220220759511
step: 140, loss: 0.05220147967338562

```

```

step: 150, loss: 0.002705523744225502
step: 160, loss: 0.0016744050662964582
step: 170, loss: 0.07002072036266327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.7990074441687345, f1=0.7915690866510539,
best_f1=0.7915690866510539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005848862696439028
step: 10, loss: 0.06383176892995834
step: 20, loss: 0.04788107052445412
step: 30, loss: 0.003919882234185934
step: 40, loss: 0.0007653604843653738
step: 50, loss: 0.0030465542804449797
step: 60, loss: 0.07095219939947128
step: 70, loss: 0.041501712054014206
step: 80, loss: 0.03489028289914131
step: 90, loss: 0.0012030103243887424
step: 100, loss: 0.0004264457384124398
step: 110, loss: 0.05349091812968254
step: 120, loss: 0.08693986386060715
step: 130, loss: 0.007932609878480434
step: 140, loss: 0.03605453670024872
step: 150, loss: 0.1739102602005005
step: 160, loss: 0.011539461091160774
step: 170, loss: 0.01078730821609497
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7810945273631842, f1=0.7728337236533958,
best_f1=0.7915690866510539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03429504483938217
step: 10, loss: 0.06597461551427841
step: 20, loss: 0.007578050717711449
step: 30, loss: 0.002947364468127489
step: 40, loss: 0.006981608457863331
step: 50, loss: 0.0046581365168094635
step: 60, loss: 0.004960197024047375

```

```

step: 70, loss: 0.005817505065351725
step: 80, loss: 0.004339949227869511
step: 90, loss: 0.0011991729261353612
step: 100, loss: 0.0010640117106959224
step: 110, loss: 0.092748261988163
step: 120, loss: 0.0020426542032510042
step: 130, loss: 0.026765041053295135
step: 140, loss: 0.0009658216731622815
step: 150, loss: 0.0008895468199625611
step: 160, loss: 0.003047303995117545
step: 170, loss: 0.003482608823105693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7980049875311721, f1=0.7865707434052758,
best_f1=0.7915690866510539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007138312794268131
step: 10, loss: 0.0008102034917101264
step: 20, loss: 0.012860089540481567
step: 30, loss: 0.13340318202972412
step: 40, loss: 0.00249027949757874
step: 50, loss: 0.20676250755786896
step: 60, loss: 0.0012358357198536396
step: 70, loss: 0.007063585799187422
step: 80, loss: 0.0036413599736988544
step: 90, loss: 0.0018261345103383064
step: 100, loss: 0.0019880738109350204
step: 110, loss: 0.006287990603595972
step: 120, loss: 0.028342289850115776
step: 130, loss: 0.003358252812176943
step: 140, loss: 0.00200459873303771
step: 150, loss: 0.00822804681956768
step: 160, loss: 0.01926468126475811
step: 170, loss: 0.0037283082492649555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7884615384615384, f1=0.7891156462585033,
best_f1=0.7915690866510539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010913343168795109
step: 10, loss: 0.008657298050820827
step: 20, loss: 0.03141740709543228

```

```

step: 30, loss: 0.11552523821592331
step: 40, loss: 0.0012262098025530577
step: 50, loss: 0.0011221661698073149
step: 60, loss: 0.0014007269637659192
step: 70, loss: 0.006213122047483921
step: 80, loss: 0.001125963288359344
step: 90, loss: 0.0014660614542663097
step: 100, loss: 0.0010843919590115547
step: 110, loss: 0.00035465534892864525
step: 120, loss: 0.06637334078550339
step: 130, loss: 0.004815959371626377
step: 140, loss: 0.012082492001354694
step: 150, loss: 0.015364422462880611
step: 160, loss: 0.0702279731631279
step: 170, loss: 0.005735345650464296
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7903614457831325, f1=0.7909090909090909,
best_f1=0.7915690866510539
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```

packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
1916it [00:05, 320.13it/s]
load_f1 = 0.6624203821656051
real_f1 = 0.6463414634146342
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 247.37it/s]

```

## 4.3 DITTO DIRTY

### 4.3.1 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize \
  --da entry_swap \
  --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \

```

```
--checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : O2
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Warning: multi\_tensor\_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda\_ext --cpp\_ext. Using Python fallback.

Original ImportError was: ModuleNotFoundError("No module named 'amp\_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/\_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0

/usr/local/lib/python3.7/dist-packages/torch/optim/lr\_scheduler.py:136:

UserWarning: Detected call of `lr\_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr\_scheduler.step()`. Failure to do this will result

in PyTorch skipping the first value of the learning rate schedule. See more

details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>", UserWarning)

```
step: 0, loss: 0.7941242456436157
step: 10, loss: 0.49005433917045593
step: 20, loss: 0.5715735554695129
step: 30, loss: 0.4096553325653076
step: 40, loss: 0.15678562223911285
step: 50, loss: 0.17223864793777466
step: 60, loss: 0.0682372972369194
step: 70, loss: 0.21719157695770264
step: 80, loss: 0.19501501321792603
```

```

step: 90, loss: 0.041025310754776
step: 100, loss: 0.15930888056755066
step: 110, loss: 0.11747141182422638
step: 120, loss: 0.04205765575170517
step: 130, loss: 0.021420452743768692
step: 140, loss: 0.015937285497784615
step: 150, loss: 0.15079304575920105
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.0509822778403759
step: 170, loss: 0.043598514050245285
step: 180, loss: 0.014789336360991001
step: 190, loss: 0.037567559629678726
step: 200, loss: 0.005879335105419159
step: 210, loss: 0.005445767194032669
step: 220, loss: 0.01671842485666275
step: 230, loss: 0.011991876177489758
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9642058165548099, f1=0.961625282167043,
best_f1=0.961625282167043
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04470125585794449
step: 10, loss: 0.021591586992144585
step: 20, loss: 0.007205067668110132
step: 30, loss: 0.008484849706292152
step: 40, loss: 0.009087535552680492
step: 50, loss: 0.011635631322860718
step: 60, loss: 0.009325980208814144
step: 70, loss: 0.009090966545045376
step: 80, loss: 0.0031185925472527742
step: 90, loss: 0.016294989734888077
step: 100, loss: 0.12891782820224762
step: 110, loss: 0.1425471305847168
step: 120, loss: 0.0025098284240812063
step: 130, loss: 0.005885434336960316
step: 140, loss: 0.13535624742507935
step: 150, loss: 0.012214804999530315
step: 160, loss: 0.0103977732360363
step: 170, loss: 0.10746699571609497
step: 180, loss: 0.022671252489089966
step: 190, loss: 0.120514415204525

```



```

step: 200, loss: 0.004278390668332577
step: 210, loss: 0.05449816212058067
step: 220, loss: 0.0015909022185951471
step: 230, loss: 0.03100520931184292
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9662921348314607, f1=0.9573991031390134,
best_f1=0.9573991031390134
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04248080775141716
step: 10, loss: 0.050099506974220276
step: 20, loss: 0.011403782293200493
step: 30, loss: 0.045286692678928375
step: 40, loss: 0.01263802032917738
step: 50, loss: 0.013107429258525372
step: 60, loss: 0.027691707015037537
step: 70, loss: 0.00251839985139668
step: 80, loss: 0.14051693677902222
step: 90, loss: 0.04750063270330429
step: 100, loss: 0.023529870435595512
step: 110, loss: 0.0374758243560791
step: 120, loss: 0.03448348492383957
step: 130, loss: 0.0022666454315185547
step: 140, loss: 0.00169152591843158
step: 150, loss: 0.0044085062108933926
step: 160, loss: 0.0037072289269417524
step: 170, loss: 0.055624812841415405
step: 180, loss: 0.01911254972219467
step: 190, loss: 0.003934381064027548
step: 200, loss: 0.003369256854057312
step: 210, loss: 0.019593728706240654
step: 220, loss: 0.0044845701195299625
step: 230, loss: 0.08034664392471313
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9696287964004499, f1=0.967305524239008,

```

```

best_f1=0.967305524239008
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005755159072577953
step: 10, loss: 0.09507782012224197
step: 20, loss: 0.09847977757453918
step: 30, loss: 0.0023577972315251827
step: 40, loss: 0.00714118592441082
step: 50, loss: 0.004877576604485512
step: 60, loss: 0.005116848275065422
step: 70, loss: 0.0488380528986454
step: 80, loss: 0.057227130979299545
step: 90, loss: 0.1280336081981659
step: 100, loss: 0.008303612470626831
step: 110, loss: 0.017189200967550278
step: 120, loss: 0.03711928427219391
step: 130, loss: 0.03389161080121994
step: 140, loss: 0.0005901542026549578
step: 150, loss: 0.0020689417142421007
step: 160, loss: 0.00666722422465682
step: 170, loss: 0.0027982336468994617
step: 180, loss: 0.061138905584812164
step: 190, loss: 0.002015385078266263
step: 200, loss: 0.029723918065428734
step: 210, loss: 0.004420800134539604
step: 220, loss: 0.006936267018318176
step: 230, loss: 0.0006351065821945667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9753914988814317, f1=0.9642857142857144,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001365576172247529
step: 10, loss: 0.008713683113455772
step: 20, loss: 0.001155296340584755
step: 30, loss: 0.001177414902485907
step: 40, loss: 0.010402112267911434
step: 50, loss: 0.0031748900655657053
step: 60, loss: 0.013570195995271206
step: 70, loss: 0.05543176829814911

```

```

step: 80, loss: 0.025699550285935402
step: 90, loss: 0.0006647765403613448
step: 100, loss: 0.0030629578977823257
step: 110, loss: 0.010822871699929237
step: 120, loss: 0.03138114884495735
step: 130, loss: 0.006124537438154221
step: 140, loss: 0.0017440543742850423
step: 150, loss: 0.000814888917375356
step: 160, loss: 0.10772479325532913
step: 170, loss: 0.06354764848947525
step: 180, loss: 0.0023813818115741014
step: 190, loss: 0.016972150653600693
step: 200, loss: 0.022602709010243416
step: 210, loss: 0.005594191607087851
step: 220, loss: 0.004290115088224411
step: 230, loss: 0.0069167171604931355
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.971687429218573, f1=0.963718820861678,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020583560690283775
step: 10, loss: 0.001216034754179418
step: 20, loss: 0.0006502160103991628
step: 30, loss: 0.03553376346826553
step: 40, loss: 0.004324525129050016
step: 50, loss: 0.003297296119853854
step: 60, loss: 0.005717396270483732
step: 70, loss: 0.25190499424934387
step: 80, loss: 0.06074999272823334
step: 90, loss: 0.0032161420676857233
step: 100, loss: 0.002714714268222451
step: 110, loss: 0.03599013760685921
step: 120, loss: 0.0005108996410854161
step: 130, loss: 0.0003020143776666373
step: 140, loss: 0.0011889231391251087
step: 150, loss: 0.0006919330917298794
step: 160, loss: 0.019993919879198074
step: 170, loss: 0.003373869229108095
step: 180, loss: 0.0006536705186590552
step: 190, loss: 0.007087760139256716
step: 200, loss: 0.157533198595047
step: 210, loss: 0.10046666115522385
step: 220, loss: 0.0007222610875032842
step: 230, loss: 0.000900878687389195

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9716231555051079, f1=0.9706546275395034,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00414420198649168
step: 10, loss: 0.006017122883349657
step: 20, loss: 0.0003321452531963587
step: 30, loss: 0.0012367625022307038
step: 40, loss: 0.008385824970901012
step: 50, loss: 0.0011002452811226249
step: 60, loss: 0.05733918398618698
step: 70, loss: 0.0018105435883626342
step: 80, loss: 0.00025522601208649576
step: 90, loss: 0.002716905903071165
step: 100, loss: 0.001749651157297194
step: 110, loss: 0.015828154981136322
step: 120, loss: 0.016348496079444885
step: 130, loss: 0.010439234785735607
step: 140, loss: 0.00016856389993336052
step: 150, loss: 0.13833338022232056
step: 160, loss: 0.0004985530395060778
step: 170, loss: 0.0015735108172520995
step: 180, loss: 0.00024357708753086627
step: 190, loss: 0.0023113288916647434
step: 200, loss: 0.0002896975784096867
step: 210, loss: 0.00024344281700905412
step: 220, loss: 0.00045581109588965774
step: 230, loss: 0.0008959274273365736
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9680264608599779, f1=0.9679558011049725,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021775062195956707
step: 10, loss: 0.008619808591902256
step: 20, loss: 0.0001236421085195616
step: 30, loss: 0.00043216050835326314
step: 40, loss: 0.0008516776724718511
step: 50, loss: 0.0005942093557678163
step: 60, loss: 0.0007196748629212379
step: 70, loss: 0.0004331687814556062

```

```

step: 80, loss: 0.0009147673845291138
step: 90, loss: 0.0002051580959232524
step: 100, loss: 0.00017980783013626933
step: 110, loss: 0.0003157068567816168
step: 120, loss: 0.00035920500522479415
step: 130, loss: 0.00800353568047285
step: 140, loss: 0.00028345902683213353
step: 150, loss: 0.0003627945261541754
step: 160, loss: 0.0014462064718827605
step: 170, loss: 0.002230361569672823
step: 180, loss: 0.0013612270122393966
step: 190, loss: 0.0005053993663750589
step: 200, loss: 0.0028842610772699118
step: 210, loss: 0.0002088800974888727
step: 220, loss: 0.002515365369617939
step: 230, loss: 0.0015889143105596304
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9731543624161074, f1=0.9666666666666666,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017067060980480164
step: 10, loss: 0.00024662973009981215
step: 20, loss: 0.0018960714805871248
step: 30, loss: 0.00464237155392766
step: 40, loss: 0.0005167190684005618
step: 50, loss: 0.0018833777867257595
step: 60, loss: 0.004996039904654026
step: 70, loss: 0.018217995762825012
step: 80, loss: 0.025711946189403534
step: 90, loss: 0.0006107217632234097
step: 100, loss: 0.013689885847270489
step: 110, loss: 0.00023834097373764962
step: 120, loss: 0.03392086178064346
step: 130, loss: 7.39123861421831e-05
step: 140, loss: 0.026176301762461662
step: 150, loss: 6.468955689342692e-05
step: 160, loss: 0.0002828034048434347
step: 170, loss: 0.00020573333313222975
step: 180, loss: 0.0018175479490309954
step: 190, loss: 0.0002638364094309509
step: 200, loss: 0.00014985950838308781
step: 210, loss: 0.00022925279336050153
step: 220, loss: 0.02437080442905426
step: 230, loss: 0.07031276822090149

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9705882352941176, f1=0.9652076318742986,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013913358270656317
step: 10, loss: 0.0004660731938201934
step: 20, loss: 0.0026241042651236057
step: 30, loss: 0.012960175052285194
step: 40, loss: 0.00011184503819094971
step: 50, loss: 0.0002990615030284971
step: 60, loss: 0.023067656904459
step: 70, loss: 0.0005434355698525906
step: 80, loss: 0.0003598322218749672
step: 90, loss: 0.002390299690887332
step: 100, loss: 0.0002298816543770954
step: 110, loss: 0.0001845713413786143
step: 120, loss: 0.01955789141356945
step: 130, loss: 0.005721569061279297
step: 140, loss: 0.024702833965420723
step: 150, loss: 0.00015589366375934333
step: 160, loss: 0.00024359596136491746
step: 170, loss: 0.0005246336222626269
step: 180, loss: 0.005448867101222277
step: 190, loss: 0.0010612605838105083
step: 200, loss: 0.0001279814459849149
step: 210, loss: 0.00020060499082319438
step: 220, loss: 0.0007634583162143826
step: 230, loss: 0.00400340324267745
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9709821428571428, f1=0.9698996655518396,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021537904103752226
step: 10, loss: 0.00046462760656140745
step: 20, loss: 3.870446380460635e-05
step: 30, loss: 0.0010191048495471478
step: 40, loss: 0.016698017716407776
step: 50, loss: 0.001271815039217472
step: 60, loss: 0.0005242902552708983
step: 70, loss: 0.00011751341662602499

```

```

step: 80, loss: 0.00015944747428875417
step: 90, loss: 7.314775575650856e-05
step: 100, loss: 5.360465729609132e-05
step: 110, loss: 6.71463058097288e-05
step: 120, loss: 0.00016593119653407484
step: 130, loss: 4.7457222535740584e-05
step: 140, loss: 5.150317520019598e-05
step: 150, loss: 4.371488103060983e-05
step: 160, loss: 0.0017394190654158592
step: 170, loss: 0.04778224974870682
step: 180, loss: 0.0005657970323227346
step: 190, loss: 0.00012514764966908842
step: 200, loss: 9.840390703175217e-05
step: 210, loss: 0.00020649618818424642
step: 220, loss: 5.703435090254061e-05
step: 230, loss: 9.075324487639591e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9728506787330317, f1=0.9708520179372198,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020451023010537028
step: 10, loss: 0.05129460245370865
step: 20, loss: 0.00019092875299975276
step: 30, loss: 0.005458608735352755
step: 40, loss: 4.7849465772742406e-05
step: 50, loss: 5.8280566008761525e-05
step: 60, loss: 0.020921463146805763
step: 70, loss: 0.04315799102187157
step: 80, loss: 0.00010122574894921854
step: 90, loss: 0.00023758172756060958
step: 100, loss: 4.620282197720371e-05
step: 110, loss: 7.953041495056823e-05
step: 120, loss: 6.0977807152085006e-05
step: 130, loss: 0.0059016696177423
step: 140, loss: 3.9740152715239674e-05
step: 150, loss: 0.00023115570365916938
step: 160, loss: 0.0010158525547012687
step: 170, loss: 4.707446714746766e-05
step: 180, loss: 0.0008694717544130981
step: 190, loss: 0.0013863536296412349
step: 200, loss: 0.017406843602657318
step: 210, loss: 4.165101927355863e-05
step: 220, loss: 0.016274893656373024
step: 230, loss: 0.00011582489241845906

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9733333333333333, f1=0.9711111111111111,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015141257317736745
step: 10, loss: 0.00010592774924589321
step: 20, loss: 0.00012747864820994437
step: 30, loss: 0.019458720460534096
step: 40, loss: 0.002278661821037531
step: 50, loss: 8.028097363421693e-05
step: 60, loss: 4.285056274966337e-05
step: 70, loss: 4.6649838623125106e-05
step: 80, loss: 5.599766518571414e-05
step: 90, loss: 3.6082154110772535e-05
step: 100, loss: 4.061559957335703e-05
step: 110, loss: 4.772997272084467e-05
step: 120, loss: 4.669067857321352e-05
step: 130, loss: 0.0009453163365833461
step: 140, loss: 0.03634963557124138
step: 150, loss: 0.0194717850536108
step: 160, loss: 0.007770358119159937
step: 170, loss: 5.239271922619082e-05
step: 180, loss: 0.00045674550347030163
step: 190, loss: 6.964828935451806e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 200, loss: 4.40153744420968e-05
step: 210, loss: 5.529753980226815e-05
step: 220, loss: 0.00013686629245057702
step: 230, loss: 3.673749597510323e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9711111111111111, f1=0.9679558011049725,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002753847511485219
step: 10, loss: 2.9268981961649843e-05
step: 20, loss: 0.004347028210759163
step: 30, loss: 5.1447768782963976e-05
step: 40, loss: 2.791674342006445e-05
step: 50, loss: 0.0015699815703555942
step: 60, loss: 6.554513674927875e-05

```



```

step: 70, loss: 3.85591592930723e-05
step: 80, loss: 5.58553401788231e-05
step: 90, loss: 0.00019388727378100157
step: 100, loss: 0.002406249986961484
step: 110, loss: 0.00014983891742303967
step: 120, loss: 0.0006672405288554728
step: 130, loss: 5.737574974773452e-05
step: 140, loss: 3.143688081763685e-05
step: 150, loss: 4.09507192671299e-05
step: 160, loss: 3.241623198846355e-05
step: 170, loss: 3.329926039441489e-05
step: 180, loss: 8.168204658431932e-05
step: 190, loss: 5.830897862324491e-05
step: 200, loss: 5.48302341485396e-05
step: 210, loss: 0.0002127140760421753
step: 220, loss: 0.00032584197469986975
step: 230, loss: 2.554758066253271e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9731543624161074, f1=0.9720670391061451,
best_f1=0.9642857142857144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.416367690078914e-05
step: 10, loss: 0.0001860470074461773
step: 20, loss: 3.1418319849763066e-05
step: 30, loss: 9.932985994964838e-05
step: 40, loss: 4.389726746012457e-05
step: 50, loss: 0.0002886130241677165
step: 60, loss: 2.8773505619028583e-05
step: 70, loss: 6.036122067598626e-05
step: 80, loss: 9.50013127294369e-05
step: 90, loss: 2.8117809051764198e-05
step: 100, loss: 0.0002997988776769489
step: 110, loss: 5.2857121772831306e-05
step: 120, loss: 6.845514144515619e-05
step: 130, loss: 0.14783845841884613
step: 140, loss: 0.028754739090800285
step: 150, loss: 0.00015041386359371245
step: 160, loss: 0.13189995288848877
step: 170, loss: 3.315017238492146e-05
step: 180, loss: 0.00011568154877750203
step: 190, loss: 0.0009230828145518899
step: 200, loss: 3.823832594207488e-05
step: 210, loss: 0.012343969196081161
step: 220, loss: 0.01935086026787758

```

```

step: 230, loss: 0.00011409536091377959
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9721913236929923, f1=0.9710467706013363,
best_f1=0.9642857142857144
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 229.69it/s]
load_f1 = 0.9798206278026906
real_f1 = 0.9764837625979844
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 248.31it/s]

```

### 4.3.2 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize \
  --da entry_swap \
  --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8008736371994019
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.44374018907546997
step: 20, loss: 0.503638744354248
step: 30, loss: 0.4407169222831726
step: 40, loss: 0.3626297116279602
step: 50, loss: 0.2117101103067398
step: 60, loss: 0.21337608993053436
step: 70, loss: 0.26050183176994324
step: 80, loss: 0.14431841671466827
step: 90, loss: 0.09042112529277802
step: 100, loss: 0.4021649956703186
step: 110, loss: 0.044530194252729416
step: 120, loss: 0.05865568667650223
step: 130, loss: 0.042833324521780014
step: 140, loss: 0.13040031492710114
step: 150, loss: 0.04192601516842842
step: 160, loss: 0.15662804245948792
step: 170, loss: 0.1519622802734375
step: 180, loss: 0.10327019542455673
step: 190, loss: 0.029868874698877335
step: 200, loss: 0.16336554288864136
step: 210, loss: 0.13252469897270203

```

```

step: 220, loss: 0.09136752784252167
step: 230, loss: 0.0710417777299881
step: 240, loss: 0.21125368773937225
step: 250, loss: 0.06869743019342422
step: 260, loss: 0.026105428114533424
step: 270, loss: 0.022056641057133675
step: 280, loss: 0.21437986195087433
step: 290, loss: 0.08823332190513611
step: 300, loss: 0.14231343567371368
step: 310, loss: 0.0413636714220047
step: 320, loss: 0.17793557047843933
step: 330, loss: 0.19070452451705933
step: 340, loss: 0.2290506362915039
step: 350, loss: 0.15782374143600464
step: 360, loss: 0.06354232132434845
step: 370, loss: 0.049175143241882324
step: 380, loss: 0.14577828347682953
step: 390, loss: 0.012472988106310368
step: 400, loss: 0.034077566117048264
step: 410, loss: 0.03664117306470871
step: 420, loss: 0.026461200788617134
step: 430, loss: 0.018611593171954155
step: 440, loss: 0.061150241643190384
step: 450, loss: 0.04191114008426666
step: 460, loss: 0.14781169593334198
step: 470, loss: 0.2900681495666504
step: 480, loss: 0.27543705701828003
step: 490, loss: 0.034140996634960175
step: 500, loss: 0.07696163654327393
step: 510, loss: 0.07992904633283615
step: 520, loss: 0.0908297672867775
step: 530, loss: 0.07040547579526901
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.929840972871843, f1=0.9298000929800094,
best_f1=0.9298000929800094
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12874986231327057
step: 10, loss: 0.15189436078071594
step: 20, loss: 0.08822006732225418
step: 30, loss: 0.10800866782665253

```

step: 40, loss: 0.01751106046140194  
step: 50, loss: 0.046734850853681564  
step: 60, loss: 0.20260484516620636  
step: 70, loss: 0.07681871950626373  
step: 80, loss: 0.005063171498477459  
step: 90, loss: 0.005652613937854767  
step: 100, loss: 0.21760688722133636  
step: 110, loss: 0.03688839077949524  
step: 120, loss: 0.042305562645196915  
step: 130, loss: 0.011090468615293503  
step: 140, loss: 0.013511580415070057  
step: 150, loss: 0.12508398294448853  
step: 160, loss: 0.06607446819543839  
step: 170, loss: 0.13063114881515503  
step: 180, loss: 0.003997074440121651  
step: 190, loss: 0.08181851357221603  
step: 200, loss: 0.1092192530632019  
step: 210, loss: 0.03745898976922035  
step: 220, loss: 0.04728894680738449  
step: 230, loss: 0.12760066986083984  
step: 240, loss: 0.1628725826740265  
step: 250, loss: 0.1285732239484787  
step: 260, loss: 0.04592607915401459  
step: 270, loss: 0.12866100668907166  
step: 280, loss: 0.07973532378673553  
step: 290, loss: 0.11926435679197311  
step: 300, loss: 0.042264677584171295  
step: 310, loss: 0.05960103124380112  
step: 320, loss: 0.24033161997795105  
step: 330, loss: 0.05038289353251457  
step: 340, loss: 0.006974378600716591  
step: 350, loss: 0.0284765362739563  
step: 360, loss: 0.027342917397618294  
step: 370, loss: 0.012738586403429508  
step: 380, loss: 0.040447842329740524  
step: 390, loss: 0.03218015283346176  
step: 400, loss: 0.055890318006277084  
step: 410, loss: 0.0008756167953833938  
step: 420, loss: 0.0680006667971611  
step: 430, loss: 0.02891637571156025  
step: 440, loss: 0.010544310323894024  
step: 450, loss: 0.03705533221364021  
step: 460, loss: 0.23440177738666534  
step: 470, loss: 0.06961242854595184  
step: 480, loss: 0.17934809625148773  
step: 490, loss: 0.027813462540507317  
step: 500, loss: 0.03674294054508209  
step: 510, loss: 0.07350566983222961

```

step: 520, loss: 0.0389007106423378
step: 530, loss: 0.14545971155166626
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9382257315373896, f1=0.9311778290993071,
best_f1=0.9311778290993071
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008399460464715958
step: 10, loss: 0.10982639342546463
step: 20, loss: 0.3248118758201599
step: 30, loss: 0.1636468768119812
step: 40, loss: 0.004897300153970718
step: 50, loss: 0.020235201343894005
step: 60, loss: 0.0023303187917917967
step: 70, loss: 0.04745110124349594
step: 80, loss: 0.011042597703635693
step: 90, loss: 0.10317692160606384
step: 100, loss: 0.012480556033551693
step: 110, loss: 0.012548001483082771
step: 120, loss: 0.026454713195562363
step: 130, loss: 0.03161470964550972
step: 140, loss: 0.022568292915821075
step: 150, loss: 0.06626187264919281
step: 160, loss: 0.005771844182163477
step: 170, loss: 0.038998786360025406
step: 180, loss: 0.009249167516827583
step: 190, loss: 0.014222001656889915
step: 200, loss: 0.04580310359597206
step: 210, loss: 0.07168065011501312
step: 220, loss: 0.025881802663207054
step: 230, loss: 0.028457196429371834
step: 240, loss: 0.05717680975794792
step: 250, loss: 0.012718926183879375
step: 260, loss: 0.009859067387878895
step: 270, loss: 0.004494919441640377
step: 280, loss: 0.036935169249773026
step: 290, loss: 0.06184135004878044
step: 300, loss: 0.044129759073257446
step: 310, loss: 0.19321472942829132
step: 320, loss: 0.13606977462768555
step: 330, loss: 0.005814980249851942

```

```

step: 340, loss: 0.00625511072576046
step: 350, loss: 0.012294615618884563
step: 360, loss: 0.03089875727891922
step: 370, loss: 0.01089672464877367
step: 380, loss: 0.030351758003234863
step: 390, loss: 0.00198223814368248
step: 400, loss: 0.006733565125614405
step: 410, loss: 0.03717721253633499
step: 420, loss: 0.04438384249806404
step: 430, loss: 0.05257474258542061
step: 440, loss: 0.029811754822731018
step: 450, loss: 0.031541258096694946
step: 460, loss: 0.16886986792087555
step: 470, loss: 0.0017288459930568933
step: 480, loss: 0.006847004871815443
step: 490, loss: 0.02820083312690258
step: 500, loss: 0.15918315947055817
step: 510, loss: 0.005254813004285097
step: 520, loss: 0.009214146994054317
step: 530, loss: 0.06637931615114212
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9354691075514875, f1=0.9372156505914468,
best_f1=0.9311778290993071
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00436615664511919
step: 10, loss: 0.0061357696540653706
step: 20, loss: 0.04132130742073059
step: 30, loss: 0.004515467211604118
step: 40, loss: 0.0018027600599452853
step: 50, loss: 0.09185291826725006
step: 60, loss: 0.0011195739498361945
step: 70, loss: 0.0019854879938066006
step: 80, loss: 0.019709773361682892
step: 90, loss: 0.010150792077183723
step: 100, loss: 0.009531430900096893
step: 110, loss: 0.0034930198453366756
step: 120, loss: 0.09866060316562653
step: 130, loss: 0.02108270861208439
step: 140, loss: 0.009535136632621288
step: 150, loss: 0.0012695512268692255
step: 160, loss: 0.0013676570961251855
step: 170, loss: 0.0062101781368255615
step: 180, loss: 0.031573474407196045
step: 190, loss: 0.013256430625915527

```



```

step: 200, loss: 0.034281544387340546
step: 210, loss: 0.04993173107504845
step: 220, loss: 0.004184755031019449
step: 230, loss: 0.1632281243801117
step: 240, loss: 0.005010286346077919
step: 250, loss: 0.011618062853813171
step: 260, loss: 0.12059356272220612
step: 270, loss: 0.06793030351400375
step: 280, loss: 0.005790361203253269
step: 290, loss: 0.10765046626329422
step: 300, loss: 0.003505128202959895
step: 310, loss: 0.001354064792394638
step: 320, loss: 0.01840495876967907
step: 330, loss: 0.04125041887164116
step: 340, loss: 0.05118338391184807
step: 350, loss: 0.023660527542233467
step: 360, loss: 0.026528380811214447
step: 370, loss: 0.03686339780688286
step: 380, loss: 0.015083536505699158
step: 390, loss: 0.2715862989425659
step: 400, loss: 0.018229125067591667
step: 410, loss: 0.11755827069282532
step: 420, loss: 0.038803573697805405
step: 430, loss: 0.011075444519519806
step: 440, loss: 0.10904967039823532
step: 450, loss: 0.009395302273333073
step: 460, loss: 0.0011315057054162025
step: 470, loss: 0.002783997217193246
step: 480, loss: 0.004073555581271648
step: 490, loss: 0.01710199937224388
step: 500, loss: 0.017764423042535782
step: 510, loss: 0.07427804917097092
step: 520, loss: 0.1103079542517662
step: 530, loss: 0.006180885247886181
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.935771214252227, f1=0.9298000929800094,
best_f1=0.9311778290993071
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036316197365522385
step: 10, loss: 0.044711172580718994
step: 20, loss: 0.04632267355918884
step: 30, loss: 0.008414199575781822
step: 40, loss: 0.014892561361193657
step: 50, loss: 0.002375767333433032

```

step: 60, loss: 0.0022550979629158974  
step: 70, loss: 0.0381438285112381  
step: 80, loss: 0.0013706537429243326  
step: 90, loss: 0.002133130095899105  
step: 100, loss: 0.0016843077028170228  
step: 110, loss: 0.005165248177945614  
step: 120, loss: 0.004046928603202105  
step: 130, loss: 0.0015748431906104088  
step: 140, loss: 0.004389472305774689  
step: 150, loss: 0.0038598193787038326  
step: 160, loss: 0.042344510555267334  
step: 170, loss: 0.006966491229832172  
step: 180, loss: 0.001641774084419012  
step: 190, loss: 0.0012073864927515388  
step: 200, loss: 0.05476551875472069  
step: 210, loss: 0.0009367023012600839  
step: 220, loss: 0.0004271396901458502  
step: 230, loss: 0.0018831307534128428  
step: 240, loss: 0.007681762333959341  
step: 250, loss: 0.026699349284172058  
step: 260, loss: 0.02660468779504299  
step: 270, loss: 0.011942744255065918  
step: 280, loss: 0.07663407176733017  
step: 290, loss: 0.0008763602236285806  
step: 300, loss: 0.15771889686584473  
step: 310, loss: 0.009554149582982063  
step: 320, loss: 0.005898150149732828  
step: 330, loss: 0.009384349919855595  
step: 340, loss: 0.005656597204506397  
step: 350, loss: 0.06946967542171478  
step: 360, loss: 0.028081713244318962  
step: 370, loss: 0.001166741130873561  
step: 380, loss: 0.17971393465995789  
step: 390, loss: 0.011410883627831936  
step: 400, loss: 0.12335120141506195  
step: 410, loss: 0.0019074102165177464  
step: 420, loss: 0.009684788063168526  
step: 430, loss: 0.005626774858683348  
step: 440, loss: 0.0019641907420009375  
step: 450, loss: 0.002249152399599552  
step: 460, loss: 0.016209037974476814  
step: 470, loss: 0.022280234843492508  
step: 480, loss: 0.15130670368671417  
step: 490, loss: 0.06895910203456879  
step: 500, loss: 0.008369969204068184  
step: 510, loss: 0.012045683339238167  
step: 520, loss: 0.0011195659171789885  
step: 530, loss: 0.01001930795609951

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9387568555758683, f1=0.9396277802995915,
best_f1=0.9396277802995915
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028353945817798376
step: 10, loss: 0.030062947422266006
step: 20, loss: 0.0032070751767605543
step: 30, loss: 0.001291907043196261
step: 40, loss: 0.003215909469872713
step: 50, loss: 0.002602900378406048
step: 60, loss: 0.0006637215265072882
step: 70, loss: 0.03178933635354042
step: 80, loss: 0.004243052564561367
step: 90, loss: 0.011961925774812698
step: 100, loss: 0.049366772174835205
step: 110, loss: 0.045167770236730576
step: 120, loss: 0.004736796487122774
step: 130, loss: 0.002228017430752516
step: 140, loss: 0.001508121145889163
step: 150, loss: 0.0018367886077612638
step: 160, loss: 0.00454045832157135
step: 170, loss: 0.0008228649967350066
step: 180, loss: 0.003924784250557423
step: 190, loss: 0.013516439124941826
step: 200, loss: 0.001760265906341374
step: 210, loss: 0.11416555196046829
step: 220, loss: 0.0029219319112598896
step: 230, loss: 0.0001412005804013461
step: 240, loss: 0.0004525971307884902
step: 250, loss: 0.0013713410589843988
step: 260, loss: 0.0011426974087953568
step: 270, loss: 0.0009360677213408053
step: 280, loss: 0.07979734987020493
step: 290, loss: 0.0006424036109820008
step: 300, loss: 0.0027422457933425903
step: 310, loss: 0.00017918953381013125
step: 320, loss: 0.003990051336586475
step: 330, loss: 0.0013191662728786469
step: 340, loss: 6.571714038727805e-05
step: 350, loss: 0.09669643640518188

```

```

step: 360, loss: 0.00026597880059853196
step: 370, loss: 0.03518805280327797
step: 380, loss: 0.017146635800600052
step: 390, loss: 0.05866008996963501
step: 400, loss: 0.008747796528041363
step: 410, loss: 0.0019334414973855019
step: 420, loss: 0.0876200720667839
step: 430, loss: 0.0007611163309775293
step: 440, loss: 0.17972132563591003
step: 450, loss: 0.00027918146224692464
step: 460, loss: 0.0033690358977764845
step: 470, loss: 0.1976771354675293
step: 480, loss: 0.014014383777976036
step: 490, loss: 0.003529304638504982
step: 500, loss: 0.0031880612950772047
step: 510, loss: 0.00016073518781922758
step: 520, loss: 0.0777079239487648
step: 530, loss: 0.004123501013964415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9396471680594244, f1=0.9295904279797514,
best_f1=0.9295904279797514
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07766084372997284
step: 10, loss: 0.038999900221824646
step: 20, loss: 0.00987187772989273
step: 30, loss: 0.008904263377189636
step: 40, loss: 0.0008964698645286262
step: 50, loss: 0.0024904434103518724
step: 60, loss: 0.15417446196079254
step: 70, loss: 0.20452973246574402
step: 80, loss: 0.013564557768404484
step: 90, loss: 0.00043577729957178235
step: 100, loss: 0.004104837775230408
step: 110, loss: 0.0004815687716472894
step: 120, loss: 0.0033185803331434727
step: 130, loss: 0.0011150394566357136
step: 140, loss: 0.004938169382512569
step: 150, loss: 0.005928738508373499
step: 160, loss: 0.0015181740745902061
step: 170, loss: 0.00017551587370689958

```

```

step: 180, loss: 0.0034047728404402733
step: 190, loss: 0.00023256533313542604
step: 200, loss: 0.0017171732615679502
step: 210, loss: 0.0008892578189261258
step: 220, loss: 0.0003800920967478305
step: 230, loss: 0.0005458273226395249
step: 240, loss: 0.02502797171473503
step: 250, loss: 0.013878758996725082
step: 260, loss: 0.09363287687301636
step: 270, loss: 0.0017635721014812589
step: 280, loss: 0.03944353386759758
step: 290, loss: 0.003409274620935321
step: 300, loss: 7.652112253708765e-05
step: 310, loss: 0.00019459309987723827
step: 320, loss: 0.11069481074810028
step: 330, loss: 0.00019173863984178752
step: 340, loss: 0.008234388194978237
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 350, loss: 0.0002483172866050154
step: 360, loss: 0.00924390647560358
step: 370, loss: 0.03655065968632698
step: 380, loss: 0.010450128465890884
step: 390, loss: 0.0021392509806901217
step: 400, loss: 0.006934883538633585
step: 410, loss: 0.0006799193215556443
step: 420, loss: 0.0011399233480915427
step: 430, loss: 0.0003597719478420913
step: 440, loss: 0.00010293089871993288
step: 450, loss: 0.0015045932959765196
step: 460, loss: 0.0007213145727291703
step: 470, loss: 0.1805097460746765
step: 480, loss: 0.00046395641402341425
step: 490, loss: 0.005009532906115055
step: 500, loss: 0.0001824970677262172
step: 510, loss: 0.02665553241968155
step: 520, loss: 0.014719175174832344
step: 530, loss: 0.0004025745438411832
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9415584415584416, f1=0.9330254041570438,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005586889456026256
step: 10, loss: 0.0569845587015152
step: 20, loss: 0.012152635492384434
step: 30, loss: 0.2084144800901413
step: 40, loss: 0.004460933152586222
step: 50, loss: 0.005882755853235722
step: 60, loss: 0.000941709557082504
step: 70, loss: 0.0001590583851793781
step: 80, loss: 0.00023082477855496109
step: 90, loss: 0.0002952887152787298
step: 100, loss: 0.0003369275655131787
step: 110, loss: 0.0017308116657659411
step: 120, loss: 0.007280795834958553
step: 130, loss: 0.002108206506818533
step: 140, loss: 0.00013324209430720657
step: 150, loss: 0.022067710757255554
step: 160, loss: 0.0006220138748176396
step: 170, loss: 0.0021105806808918715
step: 180, loss: 0.002444506622850895
step: 190, loss: 0.0008225322235375643
step: 200, loss: 0.0015875602839514613
step: 210, loss: 0.00040605387766845524
step: 220, loss: 0.0015435578534379601
step: 230, loss: 0.0006802488933317363
step: 240, loss: 0.09831523895263672
step: 250, loss: 0.0015537686413154006
step: 260, loss: 0.0976393073797226
step: 270, loss: 3.996773739345372e-05
step: 280, loss: 0.0011643852340057492
step: 290, loss: 0.03145400062203407
step: 300, loss: 6.089150701882318e-05
step: 310, loss: 0.04972522333264351
step: 320, loss: 0.005688706878572702
step: 330, loss: 0.010234436951577663
step: 340, loss: 0.0007850815891288221
step: 350, loss: 0.05119377374649048
step: 360, loss: 0.003661148715764284
step: 370, loss: 0.00017517575179226696
step: 380, loss: 0.007376641966402531
step: 390, loss: 6.73926915624179e-05
step: 400, loss: 0.04535260424017906
step: 410, loss: 0.00011380751675460488
step: 420, loss: 4.118158176424913e-05
step: 430, loss: 0.00014191775699146092
step: 440, loss: 0.00044832821004092693
step: 450, loss: 0.0001039478593156673
step: 460, loss: 0.006249665282666683
```

```

step: 470, loss: 0.0001976726925931871
step: 480, loss: 0.0005450915778055787
step: 490, loss: 0.00010325831681257114
step: 500, loss: 0.020288478583097458
step: 510, loss: 0.001709322677925229
step: 520, loss: 0.00024103639589156955
step: 530, loss: 0.00021448582992888987
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9395477618827872, f1=0.9330889092575618,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004169973079115152
step: 10, loss: 0.0027620885521173477
step: 20, loss: 9.711569146020338e-05
step: 30, loss: 0.0008696541772224009
step: 40, loss: 0.009736626408994198
step: 50, loss: 7.325731712626293e-05
step: 60, loss: 5.127083932165988e-05
step: 70, loss: 0.044671446084976196
step: 80, loss: 5.923606295255013e-05
step: 90, loss: 0.0011743141803890467
step: 100, loss: 0.00019846041686832905
step: 110, loss: 0.00020680598390754312
step: 120, loss: 0.00046609019045718014
step: 130, loss: 0.011718005873262882
step: 140, loss: 0.01337569858878851
step: 150, loss: 0.0005281227640807629
step: 160, loss: 0.0003224030078854412
step: 170, loss: 0.00013651013432536274
step: 180, loss: 0.00010796722926897928
step: 190, loss: 0.16947399079799652
step: 200, loss: 0.007839630357921124
step: 210, loss: 0.00022293729125522077
step: 220, loss: 0.01624424196779728
step: 230, loss: 0.0005237037548795342
step: 240, loss: 0.000371648115105927
step: 250, loss: 0.11146025359630585
step: 260, loss: 0.011126893572509289
step: 270, loss: 0.022331928834319115
step: 280, loss: 0.0002499468391761184
step: 290, loss: 6.262017996050417e-05
step: 300, loss: 0.0009333111229352653
step: 310, loss: 0.00015145323413889855
step: 320, loss: 0.0006531992694362998

```

```

step: 330, loss: 0.0015336287906393409
step: 340, loss: 0.0001903476077131927
step: 350, loss: 0.0004679834528360516
step: 360, loss: 0.16247865557670593
step: 370, loss: 0.0007653757929801941
step: 380, loss: 0.0013186115538701415
step: 390, loss: 0.0002705507504288107
step: 400, loss: 0.04810865595936775
step: 410, loss: 0.00019470973347779363
step: 420, loss: 0.0025481763295829296
step: 430, loss: 0.00018304397235624492
step: 440, loss: 0.0012286931741982698
step: 450, loss: 0.0002265489601995796
step: 460, loss: 0.01104152575135231
step: 470, loss: 4.105022890144028e-05
step: 480, loss: 0.0002092366776196286
step: 490, loss: 0.0005800577928312123
step: 500, loss: 0.04832042008638382
step: 510, loss: 0.001134670339524746
step: 520, loss: 3.1783394661033526e-05
step: 530, loss: 0.00591306434944272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9397477814105558, f1=0.933271547729379,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06720274686813354
step: 10, loss: 0.0003415568789932877
step: 20, loss: 0.0005560606368817389
step: 30, loss: 0.0009921470191329718
step: 40, loss: 0.03230421245098114
step: 50, loss: 0.0008389758877456188
step: 60, loss: 0.003484006505459547
step: 70, loss: 0.010522250086069107
step: 80, loss: 0.008020628243684769
step: 90, loss: 0.00030621819314546883
step: 100, loss: 0.000717099872417748
step: 110, loss: 0.0007365212659351528
step: 120, loss: 0.00046347291208803654
step: 130, loss: 6.785775622120127e-05
step: 140, loss: 0.0019314475357532501
step: 150, loss: 0.0002758621994871646
step: 160, loss: 0.000607712019700557
step: 170, loss: 0.0005674294661730528
step: 180, loss: 0.001574129331856966

```



```

step: 190, loss: 0.004787868354469538
step: 200, loss: 9.226874681189656e-05
step: 210, loss: 3.910623490810394e-05
step: 220, loss: 0.00044518039794638753
step: 230, loss: 0.005215841811150312
step: 240, loss: 0.0002616971032693982
step: 250, loss: 0.0006086459034122527
step: 260, loss: 0.006654716096818447
step: 270, loss: 0.01675296016037464
step: 280, loss: 0.0010440361220389605
step: 290, loss: 0.0019057139288634062
step: 300, loss: 0.030248919501900673
step: 310, loss: 7.375473069259897e-05
step: 320, loss: 0.00019830999372061342
step: 330, loss: 0.0004838038294110447
step: 340, loss: 0.00018393289064988494
step: 350, loss: 0.0004867013485636562
step: 360, loss: 0.004272999707609415
step: 370, loss: 0.01008281297981739
step: 380, loss: 0.002557185711339116
step: 390, loss: 0.0008485176367685199
step: 400, loss: 0.006040285807102919
step: 410, loss: 9.163249342236668e-05
step: 420, loss: 0.0005713521386496723
step: 430, loss: 0.0012428538175299764
step: 440, loss: 0.0001487098343204707
step: 450, loss: 0.0005369206774048507
step: 460, loss: 0.05656590312719345
step: 470, loss: 0.003221771214157343
step: 480, loss: 0.002096862066537142
step: 490, loss: 0.309471994638443
step: 500, loss: 0.0036593538243323565
step: 510, loss: 0.0005776030011475086
step: 520, loss: 0.004734376911073923
step: 530, loss: 0.0005969071644358337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9374407582938389, f1=0.9353327085285849,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003154908772557974
step: 10, loss: 0.02337174490094185
step: 20, loss: 0.00016657607920933515
step: 30, loss: 0.0014548114268109202
step: 40, loss: 0.0015746065182611346

```

step: 50, loss: 0.00042338293860666454  
step: 60, loss: 0.0008882951806299388  
step: 70, loss: 7.280477439053357e-05  
step: 80, loss: 0.0001486867549829185  
step: 90, loss: 0.001573400106281042  
step: 100, loss: 0.08252149820327759  
step: 110, loss: 8.3184328104835e-05  
step: 120, loss: 0.0037264463026076555  
step: 130, loss: 0.0016065151430666447  
step: 140, loss: 0.00021264229144435376  
step: 150, loss: 0.0011671859538182616  
step: 160, loss: 0.0016075734747573733  
step: 170, loss: 0.13978981971740723  
step: 180, loss: 0.0002820078516378999  
step: 190, loss: 0.010457218624651432  
step: 200, loss: 0.004661006387323141  
step: 210, loss: 0.00021014135563746095  
step: 220, loss: 0.0014570171479135752  
step: 230, loss: 0.0009189526317641139  
step: 240, loss: 0.0010830998653545976  
step: 250, loss: 0.002199669601395726  
step: 260, loss: 0.0001968338474398479  
step: 270, loss: 0.004481688607484102  
step: 280, loss: 0.006403814069926739  
step: 290, loss: 0.060497429221868515  
step: 300, loss: 0.0026604582089930773  
step: 310, loss: 0.029851458966732025  
step: 320, loss: 0.003492837306112051  
step: 330, loss: 0.0020111652556806803  
step: 340, loss: 0.002269924618303776  
step: 350, loss: 0.003921481315046549  
step: 360, loss: 0.0005185200716368854  
step: 370, loss: 0.005259886849671602  
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0  
step: 380, loss: 0.0005343575030565262  
step: 390, loss: 0.0010935903992503881  
step: 400, loss: 0.0001734150282572955  
step: 410, loss: 0.003699464723467827  
step: 420, loss: 0.0030943695455789566  
step: 430, loss: 0.00022549470304511487  
step: 440, loss: 0.00019471175619401038  
step: 450, loss: 4.8595415137242526e-05  
step: 460, loss: 0.002117111347615719  
step: 470, loss: 0.014627101831138134  
step: 480, loss: 0.00017913234478328377  
step: 490, loss: 0.06303458660840988  
step: 500, loss: 0.0008571670041419566  
step: 510, loss: 0.0007024696096777916

```

step: 520, loss: 5.1808005082421005e-05
step: 530, loss: 0.0003096780856139958
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.935813953488372, f1=0.9356779268857011,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01630515046417713
step: 10, loss: 0.021037982776761055
step: 20, loss: 4.480034840526059e-05
step: 30, loss: 5.9118618082720786e-05
step: 40, loss: 0.0017787188990041614
step: 50, loss: 0.005370411090552807
step: 60, loss: 0.00028086232487112284
step: 70, loss: 0.14914576709270477
step: 80, loss: 0.0015100067248567939
step: 90, loss: 0.0048218462616205215
step: 100, loss: 0.002293608384206891
step: 110, loss: 0.004908558446913958
step: 120, loss: 0.0011612140806391835
step: 130, loss: 0.001635465188883245
step: 140, loss: 8.415402407990769e-05
step: 150, loss: 0.0012215939350426197
step: 160, loss: 0.011022293008863926
step: 170, loss: 0.0007327942294068635
step: 180, loss: 0.00040536143933422863
step: 190, loss: 0.0042037987150251865
step: 200, loss: 0.004562992602586746
step: 210, loss: 0.002317088656127453
step: 220, loss: 0.0003061628667637706
step: 230, loss: 0.0020668571814894676
step: 240, loss: 0.0005461790133267641
step: 250, loss: 0.00019800216250587255
step: 260, loss: 0.00021312819444574416
step: 270, loss: 0.0022896870505064726
step: 280, loss: 2.1364086933317594e-05
step: 290, loss: 0.000966629188042134
step: 300, loss: 0.0010774618713185191
step: 310, loss: 0.001271117595024407
step: 320, loss: 0.021370835602283478
step: 330, loss: 4.314746547606774e-05
step: 340, loss: 0.0003318432136438787
step: 350, loss: 0.004996993578970432
step: 360, loss: 0.001097413245588541

```

```

step: 370, loss: 0.00935172289609909
step: 380, loss: 0.033532578498125076
step: 390, loss: 0.00014098506653681397
step: 400, loss: 0.001204805332235992
step: 410, loss: 0.0007043222431093454
step: 420, loss: 0.0007245564484037459
step: 430, loss: 0.00020115065854042768
step: 440, loss: 0.0001600156829226762
step: 450, loss: 0.0005111057544127107
step: 460, loss: 0.0004438869364093989
step: 470, loss: 0.0015812285710126162
step: 480, loss: 0.0002739260671660304
step: 490, loss: 0.012059173546731472
step: 500, loss: 0.0009728284203447402
step: 510, loss: 0.004671628586947918
step: 520, loss: 0.0010764570906758308
step: 530, loss: 0.0001617220841581002
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9334557136301056, f1=0.9361896080218779,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005587396444752812
step: 10, loss: 3.5250006476417184e-05
step: 20, loss: 0.012073226273059845
step: 30, loss: 0.008393760770559311
step: 40, loss: 0.0008265624637715518
step: 50, loss: 3.8691647205268964e-05
step: 60, loss: 5.772261647507548e-05
step: 70, loss: 0.000247183779720217
step: 80, loss: 0.00015649039414711297
step: 90, loss: 0.015291751362383366
step: 100, loss: 1.5772706319694407e-05
step: 110, loss: 0.0005372028681449592
step: 120, loss: 0.00014684411871712655
step: 130, loss: 0.001357804867438972
step: 140, loss: 0.0004229746700730175
step: 150, loss: 8.894897473510355e-05
step: 160, loss: 0.0019023393979296088
step: 170, loss: 5.477747617987916e-05
step: 180, loss: 5.608380160992965e-05
step: 190, loss: 0.0002580989385023713
step: 200, loss: 0.007272185757756233
step: 210, loss: 0.0010288411285728216
step: 220, loss: 0.00012643482477869838

```

```

step: 230, loss: 0.0001583792909514159
step: 240, loss: 0.0002066089800791815
step: 250, loss: 0.03118346631526947
step: 260, loss: 0.00018494723190087825
step: 270, loss: 0.0005459391977638006
step: 280, loss: 0.0010553677566349506
step: 290, loss: 0.0003206914698239416
step: 300, loss: 0.03986675664782524
step: 310, loss: 0.02859550714492798
step: 320, loss: 0.0002028286107815802
step: 330, loss: 0.00020327811944298446
step: 340, loss: 4.305030597606674e-05
step: 350, loss: 0.0009167070384137332
step: 360, loss: 0.00011209654621779919
step: 370, loss: 0.0006598603795282543
step: 380, loss: 0.006402802653610706
step: 390, loss: 0.00013337039854377508
step: 400, loss: 0.00016826565843075514
step: 410, loss: 6.78310971125029e-05
step: 420, loss: 0.00017380269127897918
step: 430, loss: 0.00029361763154156506
step: 440, loss: 0.0006830666097812355
step: 450, loss: 0.0006924837362021208
step: 460, loss: 1.7027967260219157e-05
step: 470, loss: 0.023433992639183998
step: 480, loss: 0.0006196901667863131
step: 490, loss: 0.00011619949509622529
step: 500, loss: 0.0007928256527520716
step: 510, loss: 0.0007019334007054567
step: 520, loss: 3.9471582567784935e-05
step: 530, loss: 5.802310261060484e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9372114496768237, f1=0.9388686131386861,
best_f1=0.9330254041570438
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.656366844661534e-05
step: 10, loss: 5.00203859701287e-05
step: 20, loss: 7.246722088893875e-05
step: 30, loss: 0.00013656268129125237
step: 40, loss: 0.0017877303762361407
step: 50, loss: 0.005370236001908779
step: 60, loss: 3.106672374997288e-05
step: 70, loss: 2.38225948123727e-05
step: 80, loss: 0.0006634999881498516

```

```
step: 90, loss: 1.8417456885799766e-05
step: 100, loss: 0.016673525795340538
step: 110, loss: 4.56913658126723e-05
step: 120, loss: 0.00021352867770474404
step: 130, loss: 3.9036105590639636e-05
step: 140, loss: 0.013001675717532635
step: 150, loss: 0.00046698490041308105
step: 160, loss: 0.002309823641553521
step: 170, loss: 0.0012279929360374808
step: 180, loss: 8.3567327237688e-05
step: 190, loss: 0.00019890107796527445
step: 200, loss: 0.00031972554279491305
step: 210, loss: 0.0008320767083205283
step: 220, loss: 0.00015274960605893284
step: 230, loss: 0.000378646538592875
step: 240, loss: 0.004300781991332769
step: 250, loss: 2.8495942387962714e-05
step: 260, loss: 3.2929347071330994e-05
step: 270, loss: 0.030470289289951324
step: 280, loss: 0.000930728274397552
step: 290, loss: 7.582828402519226e-05
step: 300, loss: 0.001452903845347464
step: 310, loss: 0.026659833267331123
step: 320, loss: 0.00139886315446347
step: 330, loss: 0.0002657904115039855
step: 340, loss: 0.00017374972230754793
step: 350, loss: 0.0011177995475009084
step: 360, loss: 0.0012686008121818304
step: 370, loss: 0.0014861358795315027
step: 380, loss: 0.0005192531971260905
step: 390, loss: 0.0014827075647190213
step: 400, loss: 3.743975321413018e-05
step: 410, loss: 4.846817682846449e-05
step: 420, loss: 0.0006968296365812421
step: 430, loss: 0.00265940697863698
step: 440, loss: 2.128959931724239e-05
step: 450, loss: 0.0002563819580245763
step: 460, loss: 0.004957069642841816
step: 470, loss: 2.7099118597107008e-05
step: 480, loss: 0.0015990608371794224
step: 490, loss: 0.0006601469358429313
step: 500, loss: 9.876526746666059e-05
step: 510, loss: 0.006415811367332935
step: 520, loss: 0.0001230799243785441
step: 530, loss: 0.0005868875305168331
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
epoch 14: dev_f1=0.9359513791491351, f1=0.9332096474953617,  
best_f1=0.9330254041570438  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.0008597002597525716  
step: 10, loss: 0.00019156547205056995  
step: 20, loss: 0.00015628425171598792  
step: 30, loss: 0.00016447962843813002  
step: 40, loss: 0.00018730456940829754  
step: 50, loss: 0.00043155625462532043  
step: 60, loss: 5.4988118790788576e-05  
step: 70, loss: 0.00040764574077911675  
step: 80, loss: 0.000518942775670439  
step: 90, loss: 1.4647655916633084e-05  
step: 100, loss: 3.808311521424912e-05  
step: 110, loss: 0.017857180908322334  
step: 120, loss: 0.0021818550303578377  
step: 130, loss: 0.00017261263565160334  
step: 140, loss: 8.850591257214546e-05  
step: 150, loss: 0.0002190262166550383  
step: 160, loss: 3.417674815864302e-05  
step: 170, loss: 9.801347187021747e-05  
step: 180, loss: 3.546479274518788e-05  
step: 190, loss: 0.002609085524454713  
step: 200, loss: 0.0016444717766717076  
step: 210, loss: 0.0002829697332344949  
step: 220, loss: 8.355682803085074e-05  
step: 230, loss: 0.0004777188296429813  
step: 240, loss: 0.0006492392276413739  
step: 250, loss: 0.00023937046353239566  
step: 260, loss: 0.14178411662578583  
step: 270, loss: 0.00030309779685921967  
step: 280, loss: 0.00010511577420402318  
step: 290, loss: 0.0017199042486026883  
step: 300, loss: 0.042340707033872604  
step: 310, loss: 0.0040246061980724335  
step: 320, loss: 0.0006832028739154339  
step: 330, loss: 0.002870138268917799  
step: 340, loss: 0.002560752211138606  
step: 350, loss: 0.002408932661637664  
step: 360, loss: 0.00526084378361702  
step: 370, loss: 0.00010584861593088135  
step: 380, loss: 2.7208516257815063e-05  
step: 390, loss: 6.176759052323177e-05  
step: 400, loss: 0.0013601649552583694  
step: 410, loss: 0.0004895845195278525  
step: 420, loss: 0.0009090984822250903
```

```

step: 430, loss: 3.090862446697429e-05
step: 440, loss: 0.0007902782526798546
step: 450, loss: 0.00037486612563952804
step: 460, loss: 0.0003114745777565986
step: 470, loss: 0.002665166510269046
step: 480, loss: 0.0009006347972899675
step: 490, loss: 0.0004972743918187916
step: 500, loss: 7.185870344983414e-05
step: 510, loss: 0.00011709509271895513
step: 520, loss: 0.002424452221021056
step: 530, loss: 0.00037569686537608504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9363295880149813, f1=0.9364858599907279,
best_f1=0.9330254041570438
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:22, 254.33it/s]
load_f1 = 0.9361111111111111

```



```
real_f1 = 0.9355432780847146
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 250.17it/s]
```