

AADatasets_distilbert

August 9, 2022

0.1 Install packages

```
[1]: !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 | 3.68 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 86.6 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
    |                               | 14.5 MB 56.0 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
    |                               | 689 kB 61.1 MB/s
Collecting scipy==1.3.2
  Downloading scipy-1.3.2-cp37-cp37m-manylinux1_x86_64.whl (25.2 MB)
    |                               | 25.2 MB 1.4 MB/s
Collecting sentencepiece==0.1.85
  Downloading sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
    |                               | 1.0 MB 60.5 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 57.9 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 | 15.76 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 26.9 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 75.6 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
  | 753.4 MB 6.9 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
  | 56 kB 4.2 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
  | 884 kB 64.3 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 67.7 MB/s
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: 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: 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)
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 64.6 MB/s
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)
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)
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 srsly<1.1.0,>=0.1.0
    Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
      |                               | 184 kB 73.6 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)
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)
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 63.1 MB/s
Collecting catalogue<1.1.0,>=0.0.7
    Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
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)
Collecting sacremoses
    Downloading sacremoses-0.0.53.tar.gz (880 kB)
      |                               | 880 kB 63.4 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 57.4 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: filelock in /usr/local/lib/python3.7/dist-
packages (from transformers==3.1.0->-r requirements.txt (line 11)) (3.7.1)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (4.12.0)
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: 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: 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: 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: 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: 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=1449925
    sha256=daf0eedaeb3bf5c9a6ecb2a01af9326c1a210e59bc3857b028688c1f9662e748
    Stored in directory: /root/.cache/pip/wheels/48/8b/7f/473521e0c731c6566d631b28
    1f323842bbda9bd819eb9a3ead
  Building wheel for sacremoses (setup.py) ... done
    Created wheel for sacremoses: filename=sacremoses-0.0.53-py3-none-any.whl

```

```
size=895260
sha256=3e4a83c6c6841dd1a36b142102120aed57e8d5d25e0cd2b7eb768344bcec077f
  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.

```

```
[1]: True
```

0.2 Install fp16 optimization

```
[2]: !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 | 16.55 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-5d0ykc96
Created temporary directory: /tmp/pip-req-tracker-1wyv5k45
Initialized build tracking at /tmp/pip-req-tracker-1wyv5k45
Created build tracker: /tmp/pip-req-tracker-1wyv5k45
Entered build tracker: /tmp/pip-req-tracker-1wyv5k45
```

```

Created temporary directory: /tmp/pip-install-77wrg5wu
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-hycq2ly6
  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-1wyv5k45'
  Running setup.py (path:/tmp/pip-req-build-hycq2ly6/setup.py) egg_info for
package from file:///content/ditto/apex
  Created temporary directory: /tmp/pip-pip-egg-info-2wkxyr1s
  Running command python setup.py egg_info

  torch.__version__ = 1.4.0

  running egg_info
  creating /tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-info
  writing /tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-info/PKG-INFO
  writing dependency_links to /tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-
info/dependency_links.txt
  writing top-level names to /tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-
info/top_level.txt
  writing manifest file '/tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-
info/SOURCES.txt'
  adding license file 'LICENSE'
  writing manifest file '/tmp/pip-pip-egg-info-2wkxyr1s/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-hycq2ly6 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-1wyv5k45'
Created temporary directory: /tmp/pip-unpack-skefijjs
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-n_i6_rx1
  Building wheel for apex (setup.py) ... Destination directory:
/tmp/pip-wheel-n_i6_rx1
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = ''"/tmp/pip-req-build-hycq2ly6/setup.py''';

```



```
__file__='''/tmp/pip-req-build-hycq2ly6/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-n_i6_rx1
```

```
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/mlp
copying apex/mlp/__init__.py -> build/lib/apex/mlp
copying apex/mlp/mlp.py -> build/lib/apex/mlp
creating build/lib/apex/amp
copying apex/amp/opt.py -> build/lib/apex/amp
copying apex/amp/scaler.py -> build/lib/apex/amp
copying apex/amp/frontend.py -> build/lib/apex/amp
copying apex/amp/rnn_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/__init__.py -> build/lib/apex/amp
copying apex/amp/amp.py -> build/lib/apex/amp
copying apex/amp/compat.py -> build/lib/apex/amp
copying apex/amp/wrap.py -> build/lib/apex/amp
copying apex/amp/_initialize.py -> build/lib/apex/amp
copying apex/amp/_amp_state.py -> build/lib/apex/amp
copying apex/amp/_process_optimizer.py -> build/lib/apex/amp
copying apex/amp/handle.py -> build/lib/apex/amp
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/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/__init__.py -> build/lib/apex/parallel
    copying apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/lib/apex/parallel
    copying apex/parallel/multiproc.py -> build/lib/apex/parallel
    copying apex/parallel/sync_batchnorm.py -> build/lib/apex/parallel
    copying apex/parallel/LARC.py -> build/lib/apex/parallel
    copying apex/parallel/distributed.py -> build/lib/apex/parallel
    copying apex/parallel/sync_batchnorm_kernel.py -> build/lib/apex/parallel
    copying apex/parallel/optimized_sync_batchnorm.py -> build/lib/apex/parallel
    creating build/lib/apex/optimizers
    copying apex/optimizers/fused_sgd.py -> build/lib/apex/optimizers
    copying apex/optimizers/__init__.py -> build/lib/apex/optimizers
    copying apex/optimizers/fused_adam.py -> build/lib/apex/optimizers
    copying apex/optimizers/fused_mixed_precision_lamb.py ->
build/lib/apex/optimizers
    copying apex/optimizers/fused_adagrad.py -> build/lib/apex/optimizers
    copying apex/optimizers/fused_novograd.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/fp16_optimizer.py -> build/lib/apex/fp16_utils
    copying apex/fp16_utils/__init__.py -> build/lib/apex/fp16_utils
    copying apex/fp16_utils/fp16util.py -> build/lib/apex/fp16_utils
    copying apex/fp16_utils/loss_scaler.py -> build/lib/apex/fp16_utils
    creating build/lib/apex/transformer
    copying apex/transformer/microbatches.py -> build/lib/apex/transformer
    copying apex/transformer/parallel_state.py -> build/lib/apex/transformer
    copying apex/transformer/utils.py -> 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/log_util.py -> build/lib/apex/transformer
    creating build/lib/apex/RNN
    copying apex/RNN/cells.py -> build/lib/apex/RNN
    copying apex/RNN/__init__.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/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/torch_overrides.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/functional_overrides.py -> build/lib/apex/amp/lists
    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/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/pipeline_parallel
    copying apex/transformer/pipeline_parallel/utils.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/__init__.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/p2p_communication.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/_timers.py ->
build/lib/apex/transformer/pipeline_parallel
    creating build/lib/apex/transformer/layers
    copying apex/transformer/layers/layer_norm.py ->
build/lib/apex/transformer/layers
    copying apex/transformer/layers/__init__.py ->
build/lib/apex/transformer/layers
    creating 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/__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/data.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/layers.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/random.py ->
build/lib/apex/transformer/tensor_parallel
    creating build/lib/apex/transformer/testing
    copying apex/transformer/testing/standalone_gpt.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/__init__.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/standalone_bert.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/distributed_test_base.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/arguments.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_transformer_lm.py ->
build/lib/apex/transformer/testing
    creating build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipelining.py
-> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_with_i
nterleaving.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_pipelining_withou
t_interleaving.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/common.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
    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/halo_exchangers.py ->
build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/bottleneck.py ->
build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/test.py -> build/lib/apex/contrib/bottleneck
    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/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/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/encdec_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/__init__.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/fast_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/self_multihead_attn.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/encdec_multihead_attn_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/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/optimizers
    copying apex/contrib/optimizers/fp16_optimizer.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_sgd.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/__init__.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_adam.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/distributed_fused_adam.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/distributed_fused_lamb.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_lamb.py ->
build/lib/apex/contrib/optimizers
    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/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/sparsity
    copying apex/contrib/sparsity/permutation_lib.py ->
build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/__init__.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/layer_norm.py ->
build/lib/apex/contrib/layer_norm
    copying apex/contrib/layer_norm/__init__.py ->
build/lib/apex/contrib/layer_norm

```

```

    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_exchange_module_tests.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
    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/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/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/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py
-> 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/permutation_utilities.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
    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
    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/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

```

```

    creating 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/scaler.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/rnn_compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
    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/__init__.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/torch_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/amp.py -> build/bdist.linux-x86_64/wheel/apex/amp
    copying build/lib/apex/amp/compat.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/_initialize.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/_process_optimizer.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/__init__.py -> build/bdist.linux-x86_64/wheel/apex
    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/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/__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/multiproc.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
    copying build/lib/apex/parallel/LARC.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/sync_batchnorm_kernel.py ->
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
    creating 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/__init__.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_mixed_precision_lamb.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_novograd.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/fp16_optimizer.py ->
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/fp16util.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
    creating 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/microbatches.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
  copying build/lib/apex/transformer/parallel_state.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  copying build/lib/apex/transformer/utils.py ->
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/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
  creating build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  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/__init__.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/pipeline_parallel/schedules
  copying
build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipelining.py
-> build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_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_pipelining_without_interleaving.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/_timers.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/layer_norm.py ->
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
  creating 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/__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/data.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/layers.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
    creating 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
    copying build/lib/apex/transformer/testing/__init__.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/distributed_test_base.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/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_transformer_lm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
    copying build/lib/apex/transformer/enums.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
    copying build/lib/apex/transformer/log_util.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
    creating 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/__init__.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/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/halo_exchangers.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/test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
  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
  copying build/lib/apex/contrib/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib
  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/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/encdec_multihead_attn.py ->
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/fast_self_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_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/self_multihead_attn.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/encdec_multihead_attn_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/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/optimizers
  copying build/lib/apex/contrib/optimizers/fp16_optimizer.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/__init__.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/distributed_fused_adam.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/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    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/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/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/__init__.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
    creating
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
    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/permutation
_utilities.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/sparse_masklib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    creating 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
  copying build/lib/apex/contrib/layer_norm/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
  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_exchange_module_tests.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
  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/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/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/_autocast_utils.py ->
build/bdist.linux-x86_64/wheel/apex
  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-n_i6_rx1/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=e23e458da25cd2d718b17173f176fd7046b59bf8f89b3f488c1f6ec3c6977a55
Stored in directory: /tmp/pip-ephem-wheel-cache-5d0ykc96/wheels/6b/22/a2/16655
26ee3c3061243260fd522525eca31398e04bfa5ad7e1f
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-1wyv5k45'
/content/ditto

```

```

[3]: # 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 32.9 MB/s
Collecting awscli
  Downloading awscli-1.25.47-py3-none-any.whl (3.9 MB)
    |                               | 3.9 MB 56.8 MB/s
Requirement already satisfied: PyYAML<5.5,>=3.10 in
/usr/local/lib/python3.7/dist-packages (from awscli) (3.13)
Collecting botocore==1.27.47
  Downloading botocore-1.27.47-py3-none-any.whl (9.0 MB)
    |                               | 9.0 MB 62.1 MB/s
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Collecting docutils<0.17,>=0.10
  Downloading docutils-0.16-py2.py3-none-any.whl (548 kB)
    |                               | 548 kB 68.3 MB/s
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
    |                               | 79 kB 6.9 MB/s
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in

```

```

/usr/local/lib/python3.7/dist-packages (from botocore==1.27.47->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==1.27.47->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-1.25.47 botocore-1.27.47 colorama-0.4.4
docutils-0.16 jmespath-1.0.1 rsa-4.7.2 s3transfer-0.6.0 urllib3-1.25.4

```

```
[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: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages
(from torch==1.7.1) (1.19.2)
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

```
[5]: !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 | 17.32 MiB/s, done.
Resolving deltas: 100% (600/600), done.
Checking out files: 100% (1273/1273), done.

```

0.4 remove e move

```
[6]: !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

```

```
[7]: !rm configs.json
      !mv -n /content/ditto/IC/datasesErros/AADatasets/configs.json /content/ditto/
```

1 DA

1.1 DA STRUCTURED

1.1.1 Beer - Running the matcher

```
[8]: # 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 entry_swap \
  --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, 387kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 256kB/s]
Downloading: 100% 268M/268M [00:03<00:00, 70.2MB/s]
step: 0, loss: 0.867267370223999
epoch 1: dev_f1=0.2692307692307693, f1=0.2666666666666667,
best_f1=0.2666666666666667
step: 0, loss: 0.3767739236354828
epoch 2: dev_f1=0.2828282828282828, f1=0.28, best_f1=0.28
step: 0, loss: 0.3496069610118866
epoch 3: dev_f1=0.39999999999999997, f1=0.23999999999999996,
best_f1=0.23999999999999996
```

```

step: 0, loss: 0.3687741756439209
epoch 4: dev_f1=0.45833333333333326, f1=0.43478260869565216,
best_f1=0.43478260869565216
step: 0, loss: 0.29542386531829834
epoch 5: dev_f1=0.4137931034482759, f1=0.35714285714285715,
best_f1=0.43478260869565216
step: 0, loss: 0.3223329484462738
epoch 6: dev_f1=0.48148148148148157, f1=0.4680851063829786,
best_f1=0.4680851063829786
step: 0, loss: 0.26773110032081604
epoch 7: dev_f1=0.48148148148148157, f1=0.43478260869565216,
best_f1=0.4680851063829786
step: 0, loss: 0.4327264428138733
epoch 8: dev_f1=0.6086956521739131, f1=0.36363636363636365,
best_f1=0.36363636363636365
step: 0, loss: 0.21598783135414124
epoch 9: dev_f1=0.5238095238095237, f1=0.4761904761904762,
best_f1=0.36363636363636365
step: 0, loss: 0.25792956352233887
epoch 10: dev_f1=0.55, f1=0.55, best_f1=0.36363636363636365
step: 0, loss: 0.2230343073606491
epoch 11: dev_f1=0.5454545454545455, f1=0.55, best_f1=0.36363636363636365
step: 0, loss: 0.20958645641803741
epoch 12: dev_f1=0.6206896551724138, f1=0.5294117647058824,
best_f1=0.5294117647058824
step: 0, loss: 0.14492912590503693
epoch 13: dev_f1=0.6000000000000001, f1=0.6111111111111112,
best_f1=0.5294117647058824
step: 0, loss: 0.23481981456279755
epoch 14: dev_f1=0.6000000000000001, f1=0.6111111111111112,
best_f1=0.5294117647058824
step: 0, loss: 0.23412321507930756
epoch 15: dev_f1=0.6000000000000001, f1=0.6111111111111112,
best_f1=0.5294117647058824
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: 0.8163427114486694
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4736751616001129
step: 20, loss: 0.5761004686355591
step: 30, loss: 0.39454764127731323
step: 40, loss: 0.15855826437473297
step: 50, loss: 0.32950469851493835
step: 60, loss: 0.16393885016441345
step: 70, loss: 0.08260869979858398
step: 80, loss: 0.12807632982730865
step: 90, loss: 0.01949857734143734
step: 100, loss: 0.14510460197925568
step: 110, loss: 0.030562004074454308
step: 120, loss: 0.02376704290509224
step: 130, loss: 0.003896259469911456
step: 140, loss: 0.013648707419633865
step: 150, loss: 0.08374866098165512
step: 160, loss: 0.06547094136476517
step: 170, loss: 0.0022865666542202234
step: 180, loss: 0.0031720316037535667
step: 190, loss: 0.02401391603052616
step: 200, loss: 0.0030573413241654634
step: 210, loss: 0.005162440240383148
step: 220, loss: 0.002088463632389903
step: 230, loss: 0.004642636980861425
/usr/local/lib/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=0.9854096520763187, f1=0.9853768278965129,
best_f1=0.9853768278965129
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002274883911013603
step: 10, loss: 0.0007864679209887981
step: 20, loss: 0.0036066160537302494
step: 30, loss: 0.0014902155380696058
step: 40, loss: 0.012068529613316059
step: 50, loss: 0.003599576884880662
step: 60, loss: 0.011211770586669445
step: 70, loss: 0.03546769544482231
step: 80, loss: 0.002842341549694538
step: 90, loss: 0.0014953066129237413
step: 100, loss: 0.0056587825529277325
step: 110, loss: 0.001703340094536543
step: 120, loss: 0.00195601349696517
step: 130, loss: 0.0009304865961894393
step: 140, loss: 0.285648375749588
step: 150, loss: 0.016958583146333694
step: 160, loss: 0.004250739235430956
step: 170, loss: 0.0045957718975842
step: 180, loss: 0.008392627350986004
step: 190, loss: 0.11096624284982681
step: 200, loss: 0.012148989364504814
step: 210, loss: 0.038835711777210236
step: 220, loss: 0.000592490192502737
step: 230, loss: 0.006042585242539644
/usr/local/lib/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=0.9909502262443439, f1=0.9852440408626559,
best_f1=0.9852440408626559
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05212517827749252

```

```

step: 10, loss: 0.0007254437077790499
step: 20, loss: 0.006794735789299011
step: 30, loss: 0.003745053429156542
step: 40, loss: 0.00460919039323926
step: 50, loss: 0.0014415140030905604
step: 60, loss: 0.0004304778703954071
step: 70, loss: 0.0007114337058737874
step: 80, loss: 0.000589479343034327
step: 90, loss: 0.000707054219674319
step: 100, loss: 0.0008506311569362879
step: 110, loss: 0.0006117853336036205
step: 120, loss: 0.0010196141665801406
step: 130, loss: 0.00035410907003097236
step: 140, loss: 0.0007642224663868546
step: 150, loss: 0.00036538837593980134
step: 160, loss: 0.0009627677500247955
step: 170, loss: 0.005303623154759407
step: 180, loss: 0.0008106677560135722
step: 190, loss: 0.0008158856071531773
step: 200, loss: 0.00470181368291378
step: 210, loss: 0.020415697246789932
step: 220, loss: 0.0003991251578554511
step: 230, loss: 0.11846280843019485
/usr/local/lib/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=0.992108229988726, f1=0.9887133182844244,
best_f1=0.9887133182844244
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002842988818883896
step: 10, loss: 0.0005481900880113244
step: 20, loss: 0.0002370623406022787
step: 30, loss: 0.00037606447585858405
step: 40, loss: 0.0006615419988520443
step: 50, loss: 0.004155992530286312
step: 60, loss: 0.0007480921922251582
step: 70, loss: 0.007091759238392115
step: 80, loss: 0.05045437067747116
step: 90, loss: 0.021755952388048172
step: 100, loss: 0.005854192655533552
step: 110, loss: 0.0020567772444337606
step: 120, loss: 0.018809298053383827

```

```

step: 130, loss: 0.08720418065786362
step: 140, loss: 0.0003809048212133348
step: 150, loss: 0.0006188653642311692
step: 160, loss: 0.00022603169782087207
step: 170, loss: 0.011522764340043068
step: 180, loss: 0.11052341759204865
step: 190, loss: 0.01146623119711876
step: 200, loss: 0.00023936120851431042
step: 210, loss: 0.00015755346976220608
step: 220, loss: 0.00023711913672741503
step: 230, loss: 0.00016411417163908482
/usr/local/lib/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=0.9921259842519685, f1=0.9864559819413092,
best_f1=0.9864559819413092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017571721400599927
step: 10, loss: 0.0006441828445531428
step: 20, loss: 0.0002204061165684834
step: 30, loss: 0.00010189398744842038
step: 40, loss: 0.0003244711260776967
step: 50, loss: 0.00016371031233575195
step: 60, loss: 0.00019377046555746347
step: 70, loss: 0.00010746220505097881
step: 80, loss: 0.0003179395280312747
step: 90, loss: 0.002775338012725115
step: 100, loss: 0.0023896503262221813
step: 110, loss: 0.0003659043286461383
step: 120, loss: 0.04299817234277725
step: 130, loss: 0.03668953478336334
step: 140, loss: 9.202290675602853e-05
step: 150, loss: 0.00011792820441769436
step: 160, loss: 0.005388266406953335
step: 170, loss: 0.06311062723398209
step: 180, loss: 0.0002699396281968802
step: 190, loss: 0.0002201652678195387
step: 200, loss: 0.004709696862846613
step: 210, loss: 0.00021218914480414242
step: 220, loss: 0.00012782824342139065
step: 230, loss: 0.001529879868030548
/usr/local/lib/python3.7/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=0.992108229988726, f1=0.9852774631936579,
best_f1=0.9864559819413092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004264462739229202
step: 10, loss: 0.00013508678239304572
step: 20, loss: 0.0005680983886122704
step: 30, loss: 0.0002929062466137111
step: 40, loss: 0.00015301411622203887
step: 50, loss: 0.08544831722974777
step: 60, loss: 0.036580976098775864
step: 70, loss: 0.003627959406003356
step: 80, loss: 0.003082713345065713
step: 90, loss: 0.0016949696000665426
step: 100, loss: 0.0002575983526185155
step: 110, loss: 0.04074786603450775
step: 120, loss: 0.00012782927660737187
step: 130, loss: 0.0030682338401675224
step: 140, loss: 0.0006071878597140312
step: 150, loss: 0.00025184426340274513
step: 160, loss: 0.004137490410357714
step: 170, loss: 0.0003481120220385492
step: 180, loss: 0.000564131245482713
step: 190, loss: 0.005798097234219313
step: 200, loss: 0.0004557221254799515
step: 210, loss: 0.00011406988778617233
step: 220, loss: 0.0003195889003109187
step: 230, loss: 0.002946971682831645
/usr/local/lib/python3.7/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=0.9898762654668166, f1=0.9865168539325843,
best_f1=0.9864559819413092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029692232608795166
step: 10, loss: 9.177594620268792e-05
step: 20, loss: 0.0001223307044710964
step: 30, loss: 0.00014588348858524114
step: 40, loss: 0.0007403127383440733
step: 50, loss: 9.279544610762969e-05
step: 60, loss: 0.0015034646494314075
step: 70, loss: 0.0012225862592458725
step: 80, loss: 5.929893086431548e-05

```

```

step: 90, loss: 0.002599165542051196
step: 100, loss: 0.0006846157484687865
step: 110, loss: 0.005561919417232275
step: 120, loss: 0.00025864806957542896
step: 130, loss: 0.011297953315079212
step: 140, loss: 0.00014609639765694737
step: 150, loss: 0.0005382169620133936
step: 160, loss: 0.07519399374723434
step: 170, loss: 0.00018603970238473266
step: 180, loss: 0.001057072775438428
step: 190, loss: 0.0011988559272140265
step: 200, loss: 0.0003456250415183604
step: 210, loss: 9.478275023866445e-05
step: 220, loss: 0.0004576963256113231
step: 230, loss: 0.02876180037856102
/usr/local/lib/python3.7/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=0.9899216125419933, f1=0.9854423292273236,
best_f1=0.9864559819413092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03148983418941498
step: 10, loss: 0.019986089318990707
step: 20, loss: 0.0007465931121259928
step: 30, loss: 0.0011416648048907518
step: 40, loss: 0.0005096286186017096
step: 50, loss: 0.0002785708929877728
step: 60, loss: 7.720691792201251e-05
step: 70, loss: 7.746044138912112e-05
step: 80, loss: 9.340092947240919e-05
step: 90, loss: 7.468839612556621e-05
step: 100, loss: 9.869488712865859e-05
step: 110, loss: 8.954962686402723e-05
step: 120, loss: 4.3684496631613e-05
step: 130, loss: 0.003016012255102396
step: 140, loss: 5.5618587794015184e-05
step: 150, loss: 6.78092910675332e-05
step: 160, loss: 6.868928176118061e-05
step: 170, loss: 0.00028766400646418333
step: 180, loss: 0.00120806903578341
step: 190, loss: 0.0016398512525483966
step: 200, loss: 0.015553371980786324
step: 210, loss: 5.040371615905315e-05
step: 220, loss: 7.016967720119283e-05
step: 230, loss: 0.02061610296368599
/usr/local/lib/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=0.9921436588103255, f1=0.9887640449438202,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.524554141331464e-05
step: 10, loss: 0.0001018992843455635
step: 20, loss: 0.00012020659050904214
step: 30, loss: 5.551377762458287e-05
step: 40, loss: 3.056866262340918e-05
step: 50, loss: 3.443130844971165e-05
step: 60, loss: 0.0007097740890458226
step: 70, loss: 7.756552076898515e-05
step: 80, loss: 0.037410344928503036
step: 90, loss: 0.0013277154648676515
step: 100, loss: 0.0006849212804809213
step: 110, loss: 4.2145857150899246e-05
step: 120, loss: 0.03403380885720253
step: 130, loss: 5.4165713663678616e-05
step: 140, loss: 0.00010234927322017029
step: 150, loss: 3.871530861943029e-05
step: 160, loss: 5.478409002535045e-05
step: 170, loss: 3.356696106493473e-05
step: 180, loss: 8.263054041890427e-05
step: 190, loss: 3.151507189613767e-05
step: 200, loss: 5.012052497477271e-05
step: 210, loss: 5.088748366688378e-05
step: 220, loss: 0.030076464638113976
step: 230, loss: 8.454194176010787e-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=0.9910112359550561, f1=0.9876265466816648,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.3570274758385494e-05
step: 10, loss: 3.343691787449643e-05
step: 20, loss: 4.777049980475567e-05
step: 30, loss: 5.300817792885937e-05
step: 40, loss: 2.7770853193942457e-05

```

```

step: 50, loss: 0.019412359222769737
step: 60, loss: 0.09273305535316467
step: 70, loss: 0.0002011859614867717
step: 80, loss: 0.0026800022460520267
step: 90, loss: 0.00011135465319966897
step: 100, loss: 0.00022147686104290187
step: 110, loss: 8.970469207270071e-05
step: 120, loss: 0.017855925485491753
step: 130, loss: 4.7516103222733364e-05
step: 140, loss: 0.026005225256085396
step: 150, loss: 5.064834840595722e-05
step: 160, loss: 0.00013575721823144704
step: 170, loss: 5.090237391414121e-05
step: 180, loss: 9.551112452754751e-05
step: 190, loss: 5.883589619770646e-05
step: 200, loss: 3.6859953979728743e-05
step: 210, loss: 3.791772178374231e-05
step: 220, loss: 9.964352648239583e-05
step: 230, loss: 0.00013805106573272496
/usr/local/lib/python3.7/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=0.9910112359550561, f1=0.9843400447427293,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.6628901802469045e-05
step: 10, loss: 8.961684216046706e-05
step: 20, loss: 2.2846617866889574e-05
step: 30, loss: 8.940157567849383e-05
step: 40, loss: 9.632185538066551e-05
step: 50, loss: 0.0003782989806495607
step: 60, loss: 8.952953794505447e-05
step: 70, loss: 0.0007952030864544213
step: 80, loss: 9.702183888293803e-05
step: 90, loss: 3.174942685291171e-05
step: 100, loss: 6.01527099206578e-05
step: 110, loss: 5.227505607763305e-05
step: 120, loss: 0.00498991459608078
step: 130, loss: 4.676973185269162e-05
step: 140, loss: 0.00013900059275329113
step: 150, loss: 5.770932693849318e-05
step: 160, loss: 0.0018996583530679345
step: 170, loss: 0.024785839021205902
step: 180, loss: 5.3755484259454533e-05
step: 190, loss: 0.00010096342157339677
step: 200, loss: 0.00011900567915290594

```

```

step: 210, loss: 3.6208282836014405e-05
step: 220, loss: 6.407476030290127e-05
step: 230, loss: 0.01464192382991314
/usr/local/lib/python3.7/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=0.9921259842519685, f1=0.9820224719101124,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002338881022296846
step: 10, loss: 0.007524240296334028
step: 20, loss: 3.6934179661329836e-05
step: 30, loss: 7.196723890956491e-05
step: 40, loss: 4.930181603413075e-05
step: 50, loss: 0.000179888607817702
step: 60, loss: 0.012409674935042858
step: 70, loss: 0.00010307724733138457
step: 80, loss: 3.322027259855531e-05
step: 90, loss: 8.306647214340046e-05
step: 100, loss: 3.513113551889546e-05
step: 110, loss: 4.432049536262639e-05
step: 120, loss: 5.10619611304719e-05
step: 130, loss: 8.907184383133426e-05
step: 140, loss: 3.244596882723272e-05
step: 150, loss: 2.1818472305312753e-05
step: 160, loss: 0.017633924260735512
step: 170, loss: 8.506074664182961e-05
step: 180, loss: 4.102732418687083e-05
step: 190, loss: 0.00013474581646732986
step: 200, loss: 0.09560345858335495
step: 210, loss: 3.725115311681293e-05
step: 220, loss: 0.12783598899841309
step: 230, loss: 4.7092169552342966e-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=0.9910313901345291, f1=0.9854423292273236,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001484634558437392
step: 10, loss: 3.468022987362929e-05
step: 20, loss: 8.380352664971724e-05
step: 30, loss: 0.025984734296798706
step: 40, loss: 0.0002484877477400005

```



```

step: 50, loss: 7.490823918487877e-05
step: 60, loss: 5.1324244850547984e-05
step: 70, loss: 5.199654333409853e-05
step: 80, loss: 0.0001586749858688563
step: 90, loss: 2.7990385206067003e-05
step: 100, loss: 5.2569790568668395e-05
step: 110, loss: 6.446202314691618e-05
step: 120, loss: 5.0801107136067e-05
step: 130, loss: 0.00010859464964596555
step: 140, loss: 5.003952537663281e-05
step: 150, loss: 0.01791917346417904
step: 160, loss: 3.507847941364162e-05
step: 170, loss: 0.0001232820941368118
step: 180, loss: 6.461247539846227e-05
step: 190, loss: 2.431419852655381e-05
step: 200, loss: 3.3804990380303934e-05
step: 210, loss: 5.890380998607725e-05
step: 220, loss: 2.032084194070194e-05
step: 230, loss: 3.223605381208472e-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=0.9910313901345291, f1=0.9854423292273236,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.4195934858871624e-05
step: 10, loss: 1.985122071346268e-05
step: 20, loss: 0.01693044602870941
step: 30, loss: 5.978180706733838e-05
step: 40, loss: 3.598974217311479e-05
step: 50, loss: 5.802120722364634e-05
step: 60, loss: 2.737557042564731e-05
step: 70, loss: 5.180143125471659e-05
step: 80, loss: 7.99372501205653e-05
step: 90, loss: 7.87852331995964e-05
step: 100, loss: 0.015251263976097107
step: 110, loss: 0.0018826790619641542
step: 120, loss: 3.1763331207912415e-05
step: 130, loss: 9.463048627367243e-05
step: 140, loss: 0.0019246882293373346
step: 150, loss: 7.899687625467777e-05
step: 160, loss: 2.3762551791151054e-05
step: 170, loss: 3.0404209610424004e-05
step: 180, loss: 5.738701293012127e-05
step: 190, loss: 0.011453942395746708
step: 200, loss: 2.588985807960853e-05

```

```

step: 210, loss: 0.016765190288424492
step: 220, loss: 0.00010848760575754568
step: 230, loss: 1.6171243260032497e-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=0.9910313901345291, f1=0.9854423292273236,
best_f1=0.9887640449438202
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.5446253857808188e-05
step: 10, loss: 4.049886410939507e-05
step: 20, loss: 3.5704772017197683e-05
step: 30, loss: 0.00016956156468950212
step: 40, loss: 6.82672398397699e-05
step: 50, loss: 0.00010463241778779775
step: 60, loss: 2.3375281671178527e-05
step: 70, loss: 3.546299922163598e-05
step: 80, loss: 0.0190662182867527
step: 90, loss: 2.147552186215762e-05
step: 100, loss: 3.837988697341643e-05
step: 110, loss: 5.0386992370476946e-05
step: 120, loss: 0.0015527105424553156
step: 130, loss: 4.3012994865421206e-05
step: 140, loss: 6.387165194610134e-05
step: 150, loss: 9.952620166586712e-05
step: 160, loss: 2.8616563213290647e-05
step: 170, loss: 2.214613778050989e-05
step: 180, loss: 3.752990596694872e-05
step: 190, loss: 0.016911648213863373
step: 200, loss: 3.5529366869013757e-05
step: 210, loss: 0.023389335721731186
step: 220, loss: 5.870899622095749e-05
step: 230, loss: 5.8899488067254424e-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=0.9910313901345291, f1=0.9854423292273236,
best_f1=0.9887640449438202
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, 240.50it/s]
load_f1 = 0.9921436588103255
real_f1 = 0.9932584269662922
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, 262.48it/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, 388kB/s]
 Downloading: 100% 232k/232k [00:00<00:00, 253kB/s]
 Downloading: 100% 268M/268M [00:03<00:00, 70.5MB/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: 0.7857775688171387
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45399507880210876
step: 20, loss: 0.486640065908432
step: 30, loss: 0.3865237832069397
step: 40, loss: 0.288954496383667
step: 50, loss: 0.19615407288074493
step: 60, loss: 0.22598804533481598
step: 70, loss: 0.09045792371034622
step: 80, loss: 0.09449368715286255
step: 90, loss: 0.1088218167424202
step: 100, loss: 0.3243481516838074
step: 110, loss: 0.03194693475961685
step: 120, loss: 0.04746090620756149
step: 130, loss: 0.01011939812451601
step: 140, loss: 0.17058531939983368
step: 150, loss: 0.02751215547323227
step: 160, loss: 0.12638899683952332
step: 170, loss: 0.0476665236055851
step: 180, loss: 0.10164865851402283
step: 190, loss: 0.019172485917806625
step: 200, loss: 0.07196070998907089
step: 210, loss: 0.0697445198893547
step: 220, loss: 0.080885149538517
step: 230, loss: 0.12428824603557587
step: 240, loss: 0.06041506305336952
step: 250, loss: 0.03457402437925339
step: 260, loss: 0.023410320281982422
step: 270, loss: 0.005151648074388504
step: 280, loss: 0.06475549936294556
step: 290, loss: 0.04608658701181412
step: 300, loss: 0.24643515050411224
step: 310, loss: 0.09396914392709732
step: 320, loss: 0.029318800196051598
step: 330, loss: 0.05298482999205589
step: 340, loss: 0.1273360401391983
step: 350, loss: 0.09928206354379654
step: 360, loss: 0.04232257604598999
step: 370, loss: 0.13393358886241913
```

```

step: 380, loss: 0.2084435671567917
step: 390, loss: 0.015636397525668144
step: 400, loss: 0.037748102098703384
step: 410, loss: 0.019864588975906372
step: 420, loss: 0.018432505428791046
step: 430, loss: 0.01509617455303669
step: 440, loss: 0.07477463781833649
step: 450, loss: 0.02083762176334858
step: 460, loss: 0.13653388619422913
step: 470, loss: 0.13305537402629852
step: 480, loss: 0.26477155089378357
step: 490, loss: 0.02860046923160553
step: 500, loss: 0.0030899732373654842
step: 510, loss: 0.03709933161735535
step: 520, loss: 0.064431332051754
step: 530, loss: 0.1833028644323349
/usr/local/lib/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=0.9410672853828307, f1=0.9402501157943491,
best_f1=0.9402501157943491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07916245609521866
step: 10, loss: 0.08471927791833878
step: 20, loss: 0.07272370904684067
step: 30, loss: 0.021381758153438568
step: 40, loss: 0.0014451247407123446
step: 50, loss: 0.10949423909187317
step: 60, loss: 0.08378631621599197
step: 70, loss: 0.1923316866159439
step: 80, loss: 0.03435388207435608
step: 90, loss: 0.003950447775423527
step: 100, loss: 0.20840385556221008
step: 110, loss: 0.06575436890125275
step: 120, loss: 0.05338900908827782
step: 130, loss: 0.020624332129955292
step: 140, loss: 0.03281611576676369
step: 150, loss: 0.08244577050209045
step: 160, loss: 0.015365967527031898
step: 170, loss: 0.08129711449146271
step: 180, loss: 0.02024584449827671
step: 190, loss: 0.00476827472448349

```

```

step: 200, loss: 0.009019396267831326
step: 210, loss: 0.006858282722532749
step: 220, loss: 0.1388673186302185
step: 230, loss: 0.015663906931877136
step: 240, loss: 0.07465527951717377
step: 250, loss: 0.008581187576055527
step: 260, loss: 0.013515211641788483
step: 270, loss: 0.20858797430992126
step: 280, loss: 0.1889611780643463
step: 290, loss: 0.08521639555692673
step: 300, loss: 0.028768088668584824
step: 310, loss: 0.033326905220746994
step: 320, loss: 0.0682561993598938
step: 330, loss: 0.021266167983412743
step: 340, loss: 0.020640166476368904
step: 350, loss: 0.031139330938458443
step: 360, loss: 0.033810537308454514
step: 370, loss: 0.001377573236823082
step: 380, loss: 0.040153294801712036
step: 390, loss: 0.06218346208333969
step: 400, loss: 0.06679584830999374
step: 410, loss: 0.0009421546710655093
step: 420, loss: 0.029968399554491043
step: 430, loss: 0.02595902979373932
step: 440, loss: 0.022376814857125282
step: 450, loss: 0.012747149914503098
step: 460, loss: 0.1832076460123062
step: 470, loss: 0.05098932981491089
step: 480, loss: 0.21460837125778198
step: 490, loss: 0.025918707251548767
step: 500, loss: 0.01882082037627697
step: 510, loss: 0.03395210579037666
step: 520, loss: 0.09006821364164352
step: 530, loss: 0.1054430603981018
/usr/local/lib/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=0.950296397628819, f1=0.9380692167577415,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03564760088920593
step: 10, loss: 0.010253406129777431

```

step: 20, loss: 0.09508586674928665
step: 30, loss: 0.2519742250442505
step: 40, loss: 0.0015997422160580754
step: 50, loss: 0.0036859852261841297
step: 60, loss: 0.0011312251444905996
step: 70, loss: 0.07385802268981934
step: 80, loss: 0.0023071090690791607
step: 90, loss: 0.004800950177013874
step: 100, loss: 0.09027592837810516
step: 110, loss: 0.0018927770433947444
step: 120, loss: 0.010345732793211937
step: 130, loss: 0.04395484924316406
step: 140, loss: 0.022338733077049255
step: 150, loss: 0.006018990650773048
step: 160, loss: 0.01000935398042202
step: 170, loss: 0.006889389827847481
step: 180, loss: 0.03701077774167061
step: 190, loss: 0.008552555926144123
step: 200, loss: 0.009440376423299313
step: 210, loss: 0.023914190009236336
step: 220, loss: 0.020738903433084488
step: 230, loss: 0.028368953615427017
step: 240, loss: 0.005431141704320908
step: 250, loss: 0.004499573726207018
step: 260, loss: 0.0007725833565928042
step: 270, loss: 0.00475674495100975
step: 280, loss: 0.004239360801875591
step: 290, loss: 0.026353981345891953
step: 300, loss: 0.07375237345695496
step: 310, loss: 0.1358572393655777
step: 320, loss: 0.10820156335830688
step: 330, loss: 0.011182356625795364
step: 340, loss: 0.0014001267263665795
step: 350, loss: 0.010870072059333324
step: 360, loss: 0.023173538967967033
step: 370, loss: 0.002437898190692067
step: 380, loss: 0.0038107307627797127
step: 390, loss: 0.039865199476480484
step: 400, loss: 0.03566240891814232
step: 410, loss: 0.00372280552983284
step: 420, loss: 0.031520795077085495
step: 430, loss: 0.012052305042743683
step: 440, loss: 0.013558114878833294
step: 450, loss: 0.04835297539830208
step: 460, loss: 0.040951360017061234
step: 470, loss: 0.0016881379997357726
step: 480, loss: 0.0008660181192681193
step: 490, loss: 0.0038880568463355303


```

step: 500, loss: 0.025457000359892845
step: 510, loss: 0.0029343359638005495
step: 520, loss: 0.011992137879133224
step: 530, loss: 0.03825877234339714
/usr/local/lib/python3.7/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=0.9431345353675451, f1=0.9444699403396054,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004977901000529528
step: 10, loss: 0.00698886439204216
step: 20, loss: 0.026152625679969788
step: 30, loss: 0.0019341417355462909
step: 40, loss: 0.000537136394996196
step: 50, loss: 0.007411836180835962
step: 60, loss: 0.00030091547523625195
step: 70, loss: 0.0004606518486980349
step: 80, loss: 0.01845555752515793
step: 90, loss: 0.0025064265355467796
step: 100, loss: 0.0032344432547688484
step: 110, loss: 0.0013238154351711273
step: 120, loss: 0.0023890838492661715
step: 130, loss: 0.007695865351706743
step: 140, loss: 0.03947880491614342
step: 150, loss: 0.0023298743180930614
step: 160, loss: 0.022390900179743767
step: 170, loss: 0.012790623120963573
step: 180, loss: 0.07359246164560318
step: 190, loss: 0.004460708238184452
step: 200, loss: 0.00035961545654572546
step: 210, loss: 0.07563437521457672
step: 220, loss: 0.03216930851340294
step: 230, loss: 0.15533867478370667
step: 240, loss: 0.0583663210272789
step: 250, loss: 0.007823130115866661
step: 260, loss: 0.13276542723178864
step: 270, loss: 0.027836568653583527
step: 280, loss: 0.0026299285236746073
step: 290, loss: 0.0062020933255553246
step: 300, loss: 0.02070963941514492
step: 310, loss: 0.0004715007671620697
step: 320, loss: 0.018584445118904114
step: 330, loss: 0.028251999989151955
step: 340, loss: 0.011730121448636055
step: 350, loss: 0.0001648890902288258

```

```

step: 360, loss: 0.0007992787286639214
step: 370, loss: 0.11701364070177078
step: 380, loss: 0.0004285646718926728
step: 390, loss: 0.007669819053262472
step: 400, loss: 0.07871324568986893
step: 410, loss: 0.011988851241767406
step: 420, loss: 0.0007817137520760298
step: 430, loss: 0.006129392422735691
step: 440, loss: 0.04969191551208496
step: 450, loss: 0.01335518341511488
step: 460, loss: 0.0002214192209066823
step: 470, loss: 0.02063777856528759
step: 480, loss: 0.00036313696182332933
step: 490, loss: 0.05384125933051109
step: 500, loss: 0.003940016962587833
step: 510, loss: 0.022415123879909515
step: 520, loss: 0.009464608505368233
step: 530, loss: 0.002661371836438775
/usr/local/lib/python3.7/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=0.9454209065679925, f1=0.9411764705882353,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027668278198689222
step: 10, loss: 0.006422039121389389
step: 20, loss: 0.0020327961537986994
step: 30, loss: 0.0008726969244889915
step: 40, loss: 0.023664362728595734
step: 50, loss: 0.0034288272727280855
step: 60, loss: 0.0023111021146178246
step: 70, loss: 0.0019224470015615225
step: 80, loss: 0.0015814744401723146
step: 90, loss: 0.09711897373199463
step: 100, loss: 0.00029608546174131334
step: 110, loss: 0.00833996944129467
step: 120, loss: 0.0058051347732543945
step: 130, loss: 0.0013669570907950401
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 140, loss: 0.09756357222795486
step: 150, loss: 0.0013638887321576476
step: 160, loss: 0.02183416858315468
step: 170, loss: 0.019687479361891747
step: 180, loss: 0.001652575097978115
step: 190, loss: 0.00012801609409507364
step: 200, loss: 0.0006413973751477897

```

```

step: 210, loss: 0.0011279062600806355
step: 220, loss: 7.921945507405326e-05
step: 230, loss: 0.000791469297837466
step: 240, loss: 0.007864042185246944
step: 250, loss: 0.00045546412002295256
step: 260, loss: 0.03200327232480049
step: 270, loss: 0.012098738923668861
step: 280, loss: 0.05543467029929161
step: 290, loss: 0.00034238360240124166
step: 300, loss: 0.004947102628648281
step: 310, loss: 0.0014387735864147544
step: 320, loss: 0.02039215713739395
step: 330, loss: 0.00423350790515542
step: 340, loss: 0.0159425288438797
step: 350, loss: 0.004359456244856119
step: 360, loss: 0.022952910512685776
step: 370, loss: 0.0010531817097216845
step: 380, loss: 0.02811337262392044
step: 390, loss: 0.0019908335525542498
step: 400, loss: 0.00031210266752168536
step: 410, loss: 0.00017676300194580108
step: 420, loss: 0.005029086489230394
step: 430, loss: 0.014935491606593132
step: 440, loss: 0.006339464336633682
step: 450, loss: 0.05489513278007507
step: 460, loss: 0.0006908365758135915
step: 470, loss: 0.0012495070695877075
step: 480, loss: 0.001975146122276783
step: 490, loss: 0.04149634763598442
step: 500, loss: 0.015477376990020275
step: 510, loss: 0.02484089508652687
step: 520, loss: 0.0006249496364034712
step: 530, loss: 0.011994095519185066
/usr/local/lib/python3.7/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=0.9489185457892315, f1=0.9415554532903819,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003569908731151372
step: 10, loss: 0.0006604412337765098
step: 20, loss: 0.0015562174376100302
step: 30, loss: 0.006592998281121254
step: 40, loss: 0.007184082642197609
step: 50, loss: 0.0056071812286973
step: 60, loss: 0.00048576289555057883

```

step: 70, loss: 0.007915426045656204
step: 80, loss: 0.0016058540204539895
step: 90, loss: 0.003196680685505271
step: 100, loss: 0.03173274174332619
step: 110, loss: 0.08050928264856339
step: 120, loss: 0.0012102369219064713
step: 130, loss: 0.000645261665340513
step: 140, loss: 0.0012941702734678984
step: 150, loss: 0.0004568287404254079
step: 160, loss: 0.030439196154475212
step: 170, loss: 0.0031119741033762693
step: 180, loss: 0.0003914348199032247
step: 190, loss: 0.005675595719367266
step: 200, loss: 0.0002600302395876497
step: 210, loss: 0.02137116715312004
step: 220, loss: 0.0009785311995074153
step: 230, loss: 0.00011945102596655488
step: 240, loss: 0.0008554281666874886
step: 250, loss: 0.007234618533402681
step: 260, loss: 0.0003761112748179585
step: 270, loss: 0.00019914947915822268
step: 280, loss: 0.0013730545761063695
step: 290, loss: 0.0006853540544398129
step: 300, loss: 0.0004637394449673593
step: 310, loss: 0.00032604593434371054
step: 320, loss: 0.015881240367889404
step: 330, loss: 0.0012371920747682452
step: 340, loss: 0.00024579089949838817
step: 350, loss: 0.11940273642539978
step: 360, loss: 0.0005867646541446447
step: 370, loss: 0.00022809530491940677
step: 380, loss: 0.001043142518028617
step: 390, loss: 0.0016900812042877078
step: 400, loss: 7.195328362286091e-05
step: 410, loss: 0.00021594788995571434
step: 420, loss: 0.13265924155712128
step: 430, loss: 0.033573225140571594
step: 440, loss: 0.03197897598147392
step: 450, loss: 0.021732071414589882
step: 460, loss: 0.0025445655919611454
step: 470, loss: 0.11631856113672256
step: 480, loss: 0.028602125123143196
step: 490, loss: 0.00046621126239188015
step: 500, loss: 0.009345196187496185
step: 510, loss: 0.00012063013855367899
step: 520, loss: 0.029806073755025864
step: 530, loss: 0.0001569112209836021
/usr/local/lib/python3.7/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=0.949438202247191, f1=0.940677966101695,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002180305542424321
step: 10, loss: 0.020393183454871178
step: 20, loss: 0.08375437557697296
step: 30, loss: 0.0016042647184804082
step: 40, loss: 6.748136365786195e-05
step: 50, loss: 0.0018292790045961738
step: 60, loss: 0.0002337094774702564
step: 70, loss: 0.0013087402330711484
step: 80, loss: 0.006583010777831078
step: 90, loss: 0.00024283006496261805
step: 100, loss: 0.0015400511911138892
step: 110, loss: 0.00019583466928452253
step: 120, loss: 0.00013536447659134865
step: 130, loss: 0.00014871853636577725
step: 140, loss: 0.00036562446621246636
step: 150, loss: 0.0012324599083513021
step: 160, loss: 0.0014984682202339172
step: 170, loss: 0.0066613019444048405
step: 180, loss: 0.00019195533241145313
step: 190, loss: 0.00011444528354331851
step: 200, loss: 0.0035093396436423063
step: 210, loss: 0.00016694003716111183
step: 220, loss: 0.00014804521924816072
step: 230, loss: 0.00016163750842679292
step: 240, loss: 0.0033801752142608166
step: 250, loss: 0.0006070162635296583
step: 260, loss: 0.0007127522840164602
step: 270, loss: 6.383606523741037e-05
step: 280, loss: 0.024619873613119125
step: 290, loss: 0.002343612490221858
step: 300, loss: 0.000528758333530277
step: 310, loss: 0.03897412121295929
step: 320, loss: 0.02685856819152832
step: 330, loss: 0.03134704753756523
step: 340, loss: 0.028188087046146393
step: 350, loss: 6.611390563193709e-05
step: 360, loss: 0.0049446625635027885
step: 370, loss: 0.00017339304031338543
step: 380, loss: 0.0012271992163732648
step: 390, loss: 0.00011828632705146447
step: 400, loss: 4.7743702452862635e-05
```

```

step: 410, loss: 0.0018245028331875801
step: 420, loss: 0.003469709074124694
step: 430, loss: 6.113968265708536e-05
step: 440, loss: 4.5943383156554773e-05
step: 450, loss: 6.223814125405625e-05
step: 460, loss: 0.0018871804932132363
step: 470, loss: 0.00010165979620069265
step: 480, loss: 0.011334193870425224
step: 490, loss: 6.662411760771647e-05
step: 500, loss: 3.6241704947315156e-05
step: 510, loss: 0.001647772965952754
step: 520, loss: 0.00027219633921049535
step: 530, loss: 3.7646630516974255e-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=0.946773433820066, f1=0.9380028395646002,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012269262224435806
step: 10, loss: 0.012397571466863155
step: 20, loss: 0.010788220912218094
step: 30, loss: 0.0009041738812811673
step: 40, loss: 9.134884749073535e-05
step: 50, loss: 0.006025439593940973
step: 60, loss: 0.0012817963724955916
step: 70, loss: 0.0015181063208729029
step: 80, loss: 0.00014970486517995596
step: 90, loss: 0.006830740254372358
step: 100, loss: 0.001439495594240725
step: 110, loss: 0.0008379443315789104
step: 120, loss: 0.02712024748325348
step: 130, loss: 0.00020926131401211023
step: 140, loss: 2.9324677598197013e-05
step: 150, loss: 0.0010726780164986849
step: 160, loss: 0.0004566771676763892
step: 170, loss: 3.705358903971501e-05
step: 180, loss: 0.003364261705428362
step: 190, loss: 0.0027098816353827715
step: 200, loss: 8.680555038154125e-05
step: 210, loss: 6.750508327968419e-05
step: 220, loss: 0.0008225693600252271
step: 230, loss: 0.1954968124628067
step: 240, loss: 0.00016484326624777168
step: 250, loss: 7.36267538741231e-05
step: 260, loss: 0.00038008493720553815

```

```

step: 270, loss: 3.423056477913633e-05
step: 280, loss: 0.00030631310073658824
step: 290, loss: 7.735752296866849e-05
step: 300, loss: 4.038318729726598e-05
step: 310, loss: 0.029750095680356026
step: 320, loss: 4.4133852497907355e-05
step: 330, loss: 0.004091340582817793
step: 340, loss: 0.0023331877309828997
step: 350, loss: 0.017848502844572067
step: 360, loss: 0.0001643790164962411
step: 370, loss: 0.0008673085831105709
step: 380, loss: 0.1154022365808487
step: 390, loss: 0.006465431302785873
step: 400, loss: 0.04010393097996712
step: 410, loss: 0.19670817255973816
step: 420, loss: 0.00016323445015586913
step: 430, loss: 0.01692475937306881
step: 440, loss: 0.001951279235072434
step: 450, loss: 0.0007778607541695237
step: 460, loss: 0.0011752714635804296
step: 470, loss: 0.00019642134429886937
step: 480, loss: 0.00267745996825397
step: 490, loss: 0.00027681206120178103
step: 500, loss: 0.00047093600733205676
step: 510, loss: 0.00010586142161628231
step: 520, loss: 7.864127110224217e-05
step: 530, loss: 0.00018257486226502806
/usr/local/lib/python3.7/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=0.9461862423958821, f1=0.9422263973696571,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004511374281719327
step: 10, loss: 0.00018967014329973608
step: 20, loss: 8.274503488792107e-05
step: 30, loss: 0.00012571031402330846
step: 40, loss: 0.0011019876692444086
step: 50, loss: 0.00010426316293887794
step: 60, loss: 6.64155522827059e-05
step: 70, loss: 0.10620348155498505
step: 80, loss: 0.001283149467781186
step: 90, loss: 0.007929553277790546
step: 100, loss: 8.058977255132049e-05
step: 110, loss: 0.0013987708371132612
step: 120, loss: 0.000154695357196033

```

```

step: 130, loss: 0.0020762376952916384
step: 140, loss: 0.00021098231081850827
step: 150, loss: 5.175707701710053e-05
step: 160, loss: 5.102904833620414e-05
step: 170, loss: 8.454182534478605e-05
step: 180, loss: 3.8099758967291564e-05
step: 190, loss: 0.0003404048038646579
step: 200, loss: 7.933683809824288e-05
step: 210, loss: 4.913796146865934e-05
step: 220, loss: 0.00026635860558599234
step: 230, loss: 0.00018568286031950265
step: 240, loss: 8.065294969128445e-05
step: 250, loss: 0.0017235912382602692
step: 260, loss: 0.0001048042468028143
step: 270, loss: 0.0003633972373791039
step: 280, loss: 6.105479405960068e-05
step: 290, loss: 4.050210554851219e-05
step: 300, loss: 0.00013814719568472356
step: 310, loss: 3.803909567068331e-05
step: 320, loss: 0.000819320441223681
step: 330, loss: 0.009133514016866684
step: 340, loss: 5.2866904297843575e-05
step: 350, loss: 3.820429265033454e-05
step: 360, loss: 0.028106391429901123
step: 370, loss: 0.00010153867333428934
step: 380, loss: 3.1354335078503937e-05
step: 390, loss: 9.52267728280276e-05
step: 400, loss: 0.0132747832685709
step: 410, loss: 0.00013236493396107107
step: 420, loss: 7.582945545436814e-05
step: 430, loss: 8.79861181601882e-05
step: 440, loss: 4.932410956826061e-05
step: 450, loss: 5.12026745127514e-05
step: 460, loss: 0.08311862498521805
step: 470, loss: 9.524794586468488e-05
step: 480, loss: 7.760084554320201e-05
step: 490, loss: 3.69200024579186e-05
step: 500, loss: 0.002680531470105052
step: 510, loss: 0.002866383409127593
step: 520, loss: 6.322393892332911e-05
step: 530, loss: 3.3098291169153526e-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=0.9464285714285713, f1=0.9403475810239549,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006472146138548851
step: 10, loss: 2.1081079466966912e-05
step: 20, loss: 5.736055027227849e-05
step: 30, loss: 4.029689807794057e-05
step: 40, loss: 8.334668382303789e-05
step: 50, loss: 0.00010012086568167433
step: 60, loss: 7.564284169347957e-05
step: 70, loss: 5.700395195162855e-05
step: 80, loss: 0.04871796444058418
step: 90, loss: 8.125126623781398e-05
step: 100, loss: 7.384172931779176e-05
step: 110, loss: 0.0013946070102974772
step: 120, loss: 0.00010528492566663772
step: 130, loss: 3.4937969758175313e-05
step: 140, loss: 0.0006665412802249193
step: 150, loss: 2.535719977458939e-05
step: 160, loss: 7.19599484000355e-05
step: 170, loss: 0.004226521588861942
step: 180, loss: 0.0019591343589127064
step: 190, loss: 0.0002062537387246266
step: 200, loss: 0.04449600353837013
step: 210, loss: 0.0007246562163345516
step: 220, loss: 0.0001302091113757342
step: 230, loss: 8.91752788447775e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 240, loss: 4.6307519369293004e-05
step: 250, loss: 4.137195719522424e-05
step: 260, loss: 0.00012476994015742093
step: 270, loss: 0.0003378873807378113
step: 280, loss: 4.723994788946584e-05
step: 290, loss: 9.488195064477623e-05
step: 300, loss: 0.0009289195877499878
step: 310, loss: 0.0028299186378717422
step: 320, loss: 0.0004873349389526993
step: 330, loss: 4.917751721222885e-05
step: 340, loss: 6.0756261518690735e-05
step: 350, loss: 3.1917217711452395e-05
step: 360, loss: 0.0006414586096070707
step: 370, loss: 0.00020800434867851436
step: 380, loss: 4.760996307595633e-05
step: 390, loss: 4.3453288526507095e-05
step: 400, loss: 0.013422418385744095
step: 410, loss: 0.06390949338674545
step: 420, loss: 0.011967973783612251
step: 430, loss: 5.0705588364508e-05
step: 440, loss: 4.7679608542239293e-05
step: 450, loss: 0.0002889646275434643
```

```

step: 460, loss: 0.0001064567913999781
step: 470, loss: 2.029498682531994e-05
step: 480, loss: 0.00035834102891385555
step: 490, loss: 0.05567250773310661
step: 500, loss: 0.006986421532928944
step: 510, loss: 0.0012679726351052523
step: 520, loss: 0.0019416764844208956
step: 530, loss: 6.604439113289118e-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=0.9468822170900694, f1=0.9391705069124424,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008273324929177761
step: 10, loss: 0.0028801586013287306
step: 20, loss: 0.000873856944963336
step: 30, loss: 0.025847870856523514
step: 40, loss: 0.012101667933166027
step: 50, loss: 0.014390277676284313
step: 60, loss: 0.00021562348410952836
step: 70, loss: 0.00013902700447943062
step: 80, loss: 0.0003151132259517908
step: 90, loss: 9.217860497301444e-05
step: 100, loss: 0.0002745100064203143
step: 110, loss: 7.806080247974023e-05
step: 120, loss: 0.00018896767869591713
step: 130, loss: 3.314720743219368e-05
step: 140, loss: 0.0002802075177896768
step: 150, loss: 0.00014517999079544097
step: 160, loss: 6.633988959947601e-05
step: 170, loss: 0.00034112250432372093
step: 180, loss: 2.543514892749954e-05
step: 190, loss: 0.00013470540579874068
step: 200, loss: 0.002942719031125307
step: 210, loss: 5.829110523336567e-05
step: 220, loss: 0.0005072838976047933
step: 230, loss: 3.268335422035307e-05
step: 240, loss: 6.185338861541823e-05
step: 250, loss: 0.00014626314805354923
step: 260, loss: 4.478963455767371e-05
step: 270, loss: 0.0016035708831623197
step: 280, loss: 9.546819637762383e-05
step: 290, loss: 0.0035488733556121588
step: 300, loss: 0.02597762830555439
step: 310, loss: 8.463498670607805e-05

```

```

step: 320, loss: 0.01664779521524906
step: 330, loss: 0.001646105432882905
step: 340, loss: 0.0002959872072096914
step: 350, loss: 0.003131131874397397
step: 360, loss: 5.5326512665487826e-05
step: 370, loss: 1.6916161257540807e-05
step: 380, loss: 1.6357420463464223e-05
step: 390, loss: 0.004662375431507826
step: 400, loss: 1.8257138435728848e-05
step: 410, loss: 0.0004401214828249067
step: 420, loss: 0.0001813198032323271
step: 430, loss: 0.027560926973819733
step: 440, loss: 7.50027465983294e-05
step: 450, loss: 0.00017590403149370104
step: 460, loss: 0.002241838723421097
step: 470, loss: 0.0003432481607887894
step: 480, loss: 4.067799818585627e-05
step: 490, loss: 0.022295698523521423
step: 500, loss: 4.833465573028661e-05
step: 510, loss: 4.708871347247623e-05
step: 520, loss: 5.8154662838205695e-05
step: 530, loss: 0.00042309818672947586
/usr/local/lib/python3.7/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=0.9474165523548238, f1=0.935763097949886,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001834891882026568
step: 10, loss: 7.883347279857844e-05
step: 20, loss: 1.7203112292918377e-05
step: 30, loss: 4.64239674329292e-05
step: 40, loss: 6.127814413048327e-05
step: 50, loss: 5.000259989174083e-05
step: 60, loss: 0.025321828201413155
step: 70, loss: 0.00015255076868925244
step: 80, loss: 0.00330822984687984
step: 90, loss: 0.00013325836334843189
step: 100, loss: 0.0006233708118088543
step: 110, loss: 3.106676012976095e-05
step: 120, loss: 0.00016134223551489413
step: 130, loss: 3.855768227367662e-05
step: 140, loss: 2.811747253872454e-05
step: 150, loss: 0.001092533115297556
step: 160, loss: 0.00023694564879406244
step: 170, loss: 2.7777407012763433e-05

```

```

step: 180, loss: 0.0005463280831463635
step: 190, loss: 3.019181349372957e-05
step: 200, loss: 0.00034722560667432845
step: 210, loss: 0.0018978071166202426
step: 220, loss: 2.467552258167416e-05
step: 230, loss: 0.01047864742577076
step: 240, loss: 5.125622556079179e-05
step: 250, loss: 0.00017814636521507055
step: 260, loss: 4.7188277676468715e-05
step: 270, loss: 9.002309525385499e-05
step: 280, loss: 2.3855875042499974e-05
step: 290, loss: 0.0009022441226989031
step: 300, loss: 4.192049891571514e-05
step: 310, loss: 0.00010634225327521563
step: 320, loss: 0.0001223952858708799
step: 330, loss: 2.286133349116426e-05
step: 340, loss: 7.190008909674361e-05
step: 350, loss: 0.02498733624815941
step: 360, loss: 0.0011181036243215203
step: 370, loss: 4.662440187530592e-05
step: 380, loss: 6.943787593627349e-05
step: 390, loss: 2.7458025215310045e-05
step: 400, loss: 0.0003503826737869531
step: 410, loss: 0.0021119066514074802
step: 420, loss: 0.0008705162908881903
step: 430, loss: 4.109554720344022e-05
step: 440, loss: 0.0001366606302326545
step: 450, loss: 0.003247747663408518
step: 460, loss: 0.0006511214887723327
step: 470, loss: 7.400802860502154e-05
step: 480, loss: 0.0001948988065123558
step: 490, loss: 0.010782898403704166
step: 500, loss: 0.004965747240930796
step: 510, loss: 4.2629257222870365e-05
step: 520, loss: 0.019267085939645767
step: 530, loss: 0.0005088914767839015
/usr/local/lib/python3.7/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=0.9419953596287702, f1=0.9380449747590638,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011745938536478207
step: 10, loss: 0.0013692359207198024
step: 20, loss: 0.020284894853830338
step: 30, loss: 0.0001838958851294592

```

step: 40, loss: 5.028178929933347e-05
step: 50, loss: 2.964088889712002e-05
step: 60, loss: 6.030509030097164e-05
step: 70, loss: 2.1743961042375304e-05
step: 80, loss: 2.4005146769923158e-05
step: 90, loss: 0.00010470147390151396
step: 100, loss: 3.890423977281898e-05
step: 110, loss: 4.586001159623265e-05
step: 120, loss: 5.021566539653577e-05
step: 130, loss: 0.001548765110783279
step: 140, loss: 8.171649824362248e-05
step: 150, loss: 0.00010999840014846995
step: 160, loss: 5.3721370932180434e-05
step: 170, loss: 0.005082631949335337
step: 180, loss: 1.8242399164591916e-05
step: 190, loss: 2.694803333724849e-05
step: 200, loss: 0.003288146574050188
step: 210, loss: 0.0017573680961504579
step: 220, loss: 0.0003351482446305454
step: 230, loss: 0.002172443550080061
step: 240, loss: 4.305067704990506e-05
step: 250, loss: 0.057980939745903015
step: 260, loss: 0.00021125483908690512
step: 270, loss: 0.00021854587248526514
step: 280, loss: 0.0001381582405883819
step: 290, loss: 4.9095375288743526e-05
step: 300, loss: 9.148369281319901e-05
step: 310, loss: 0.00015678191266488284
step: 320, loss: 3.8743182813050225e-05
step: 330, loss: 7.289975474122912e-05
step: 340, loss: 3.350217593833804e-05
step: 350, loss: 0.03417292982339859
step: 360, loss: 4.246315438649617e-05
step: 370, loss: 0.0005078043323010206
step: 380, loss: 0.0003016756963916123
step: 390, loss: 2.2910091502126306e-05
step: 400, loss: 1.9263101421529427e-05
step: 410, loss: 4.0509941754862666e-05
step: 420, loss: 2.0045472410856746e-05
step: 430, loss: 0.00026336131850257516
step: 440, loss: 0.00012145738583058119
step: 450, loss: 0.0019994524773210287
step: 460, loss: 4.39905961684417e-05
step: 470, loss: 0.025893928483128548
step: 480, loss: 0.0015792973572388291
step: 490, loss: 0.00012144458742113784
step: 500, loss: 2.9149563488317654e-05
step: 510, loss: 2.4627366656204686e-05

```

step: 520, loss: 3.2983698474708945e-05
step: 530, loss: 6.146425585029647e-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=0.9434137291280148, f1=0.9352319706017456,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.8199487132951617e-05
step: 10, loss: 0.00010648014722391963
step: 20, loss: 5.65812224522233e-05
step: 30, loss: 2.8009808374918066e-05
step: 40, loss: 0.0001302104938076809
step: 50, loss: 0.00014775530144106597
step: 60, loss: 3.9853257476352155e-05
step: 70, loss: 2.0805280655622482e-05
step: 80, loss: 2.0678751752711833e-05
step: 90, loss: 1.4766885215067305e-05
step: 100, loss: 4.176112270215526e-05
step: 110, loss: 9.322130790678784e-05
step: 120, loss: 3.9956907130545005e-05
step: 130, loss: 4.2576983105391264e-05
step: 140, loss: 0.00020816049072891474
step: 150, loss: 0.00011110213381471112
step: 160, loss: 0.0012698133941739798
step: 170, loss: 4.9796119128586724e-05
step: 180, loss: 1.9285513189970516e-05
step: 190, loss: 5.714502913178876e-05
step: 200, loss: 3.634124368545599e-05
step: 210, loss: 7.475245365640149e-05
step: 220, loss: 4.3241208913968876e-05
step: 230, loss: 0.0022085439413785934
step: 240, loss: 2.940260492323432e-05
step: 250, loss: 3.303078847238794e-05
step: 260, loss: 1.475570479669841e-05
step: 270, loss: 0.017077624797821045
step: 280, loss: 3.821326026809402e-05
step: 290, loss: 3.2467389246448874e-05
step: 300, loss: 1.8816075680661015e-05
step: 310, loss: 0.0005021488177590072
step: 320, loss: 0.001122033572755754
step: 330, loss: 6.520273018395528e-05
step: 340, loss: 4.0509628888685256e-05
step: 350, loss: 0.0009068642393685877
step: 360, loss: 0.0037633709143847227
step: 370, loss: 2.967756154248491e-05

```

```
step: 380, loss: 3.377921166247688e-05
step: 390, loss: 2.884296736738179e-05
step: 400, loss: 2.5379851649631746e-05
step: 410, loss: 1.7769400074030273e-05
step: 420, loss: 6.215857138158754e-05
step: 430, loss: 2.0719657186418772e-05
step: 440, loss: 1.5396355593111366e-05
step: 450, loss: 2.1684280000044964e-05
step: 460, loss: 4.409367102198303e-05
step: 470, loss: 1.8279621144756675e-05
step: 480, loss: 1.9739938579732552e-05
step: 490, loss: 2.5479885152890347e-05
step: 500, loss: 4.027882096124813e-05
step: 510, loss: 3.63107756129466e-05
step: 520, loss: 2.3155967937782407e-05
step: 530, loss: 0.00010883371578529477
/usr/local/lib/python3.7/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=0.9483870967741935, f1=0.9407985314364389,
best_f1=0.9380692167577415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.553050429327413e-05
step: 10, loss: 4.84751581097953e-05
step: 20, loss: 2.52940044447314e-05
step: 30, loss: 3.43450483342167e-05
step: 40, loss: 2.169516847061459e-05
step: 50, loss: 0.0002526837051846087
step: 60, loss: 0.0004162039258517325
step: 70, loss: 0.00024193084391299635
step: 80, loss: 2.4858252800186165e-05
step: 90, loss: 1.4878641195537057e-05
step: 100, loss: 0.00021967518841847777
step: 110, loss: 0.018694356083869934
step: 120, loss: 0.0019268336473032832
step: 130, loss: 2.0764007786056027e-05
step: 140, loss: 1.7206672055181116e-05
step: 150, loss: 0.0004521789087448269
step: 160, loss: 3.080652459175326e-05
step: 170, loss: 3.205070606782101e-05
step: 180, loss: 1.5321904356824234e-05
step: 190, loss: 0.0001328680373262614
step: 200, loss: 2.4090846636681817e-05
step: 210, loss: 2.692870475584641e-05
step: 220, loss: 1.3705187484447379e-05
step: 230, loss: 2.106961255776696e-05
```

```

step: 240, loss: 2.3479997253161855e-05
step: 250, loss: 2.768899139482528e-05
step: 260, loss: 2.9345217626541853e-05
step: 270, loss: 0.09907569736242294
step: 280, loss: 0.00027863646391779184
step: 290, loss: 0.0018613961292430758
step: 300, loss: 5.5823846196290106e-05
step: 310, loss: 8.377446647500619e-05
step: 320, loss: 2.7033664082409814e-05
step: 330, loss: 1.9278073523310013e-05
step: 340, loss: 9.067162318388e-05
step: 350, loss: 2.5059363906621e-05
step: 360, loss: 4.287298725103028e-05
step: 370, loss: 2.2890882974024862e-05
step: 380, loss: 1.9445587895461358e-05
step: 390, loss: 2.0209141439408995e-05
step: 400, loss: 4.111874659429304e-05
step: 410, loss: 0.0007332014502026141
step: 420, loss: 3.547690357663669e-05
step: 430, loss: 2.268258685944602e-05
step: 440, loss: 0.00013085630780551583
step: 450, loss: 0.0004570365126710385
step: 460, loss: 1.4707255104440264e-05
step: 470, loss: 8.56419574120082e-05
step: 480, loss: 0.0028758696280419827
step: 490, loss: 0.0007934900932013988
step: 500, loss: 0.0009655792964622378
step: 510, loss: 2.752072032308206e-05
step: 520, loss: 6.086513531045057e-05
step: 530, loss: 2.1159135940251872e-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=0.9479502533394749, f1=0.9404216315307058,
best_f1=0.9380692167577415
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, 262.15it/s]

load_f1 = 0.9455216989843028

real_f1 = 0.945387792565397

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, 257.49it/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 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: 0.8431780338287354
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06621164083480835
step: 20, loss: 0.3728555142879486
step: 30, loss: 0.3702370524406433
step: 40, loss: 0.4726085662841797
step: 50, loss: 0.2778439223766327
step: 60, loss: 0.31875455379486084
step: 70, loss: 0.1761353760957718
step: 80, loss: 0.3578784763813019
step: 90, loss: 0.3281710147857666
step: 100, loss: 0.10442746430635452
step: 110, loss: 0.3318071663379669
step: 120, loss: 0.18348075449466705
step: 130, loss: 0.21594980359077454
step: 140, loss: 0.2719886302947998
step: 150, loss: 0.2666327953338623
step: 160, loss: 0.17735397815704346
step: 170, loss: 0.06301363557577133
step: 180, loss: 0.13421350717544556
step: 190, loss: 0.17334437370300293
step: 200, loss: 0.11042601615190506
step: 210, loss: 0.30512258410453796
/usr/local/lib/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=0.7183364839319472, f1=0.7154150197628458,
best_f1=0.7154150197628458
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06321427971124649
step: 10, loss: 0.07392862439155579
step: 20, loss: 0.1546081006526947
step: 30, loss: 0.11472637206315994
step: 40, loss: 0.0662708580493927
step: 50, loss: 0.21019262075424194
step: 60, loss: 0.07468581199645996
step: 70, loss: 0.14795316755771637
step: 80, loss: 0.08197010308504105
step: 90, loss: 0.02900497056543827
step: 100, loss: 0.057609569281339645
step: 110, loss: 0.11223404109477997

```

```

step: 120, loss: 0.16849440336227417
step: 130, loss: 0.17546389997005463
step: 140, loss: 0.12475868314504623
step: 150, loss: 0.10905177891254425
step: 160, loss: 0.15486404299736023
step: 170, loss: 0.12631545960903168
step: 180, loss: 0.16343684494495392
step: 190, loss: 0.14233241975307465
step: 200, loss: 0.16235142946243286
step: 210, loss: 0.06746191531419754
/usr/local/lib/python3.7/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=0.7036247334754798, f1=0.6995708154506437,
best_f1=0.7154150197628458
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17110493779182434
step: 10, loss: 0.20585505664348602
step: 20, loss: 0.24481362104415894
step: 30, loss: 0.07332152873277664
step: 40, loss: 0.2307973951101303
step: 50, loss: 0.06414566934108734
step: 60, loss: 0.09966474026441574
step: 70, loss: 0.08486360311508179
step: 80, loss: 0.031275853514671326
step: 90, loss: 0.1045476645231247
step: 100, loss: 0.05461065098643303
step: 110, loss: 0.2227332592010498
step: 120, loss: 0.21137160062789917
step: 130, loss: 0.15427376329898834
step: 140, loss: 0.24193529784679413
step: 150, loss: 0.17138701677322388
step: 160, loss: 0.06165793165564537
step: 170, loss: 0.07201540470123291
step: 180, loss: 0.046256985515356064
step: 190, loss: 0.1049683541059494
step: 200, loss: 0.09896732121706009
step: 210, loss: 0.09949565678834915
/usr/local/lib/python3.7/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=0.717948717948718, f1=0.736, best_f1=0.7154150197628458
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18922369182109833

```

```

step: 10, loss: 0.016518842428922653
step: 20, loss: 0.05820617452263832
step: 30, loss: 0.09024029225111008
step: 40, loss: 0.09110023826360703
step: 50, loss: 0.11636525392532349
step: 60, loss: 0.031153200194239616
step: 70, loss: 0.10178839415311813
step: 80, loss: 0.17981752753257751
step: 90, loss: 0.03603960573673248
step: 100, loss: 0.05976887792348862
step: 110, loss: 0.3027242422103882
step: 120, loss: 0.02474929206073284
step: 130, loss: 0.07540766894817352
step: 140, loss: 0.1253153681755066
step: 150, loss: 0.07107427716255188
step: 160, loss: 0.07189638167619705
step: 170, loss: 0.05658639222383499
step: 180, loss: 0.20765239000320435
step: 190, loss: 0.06371194869279861
step: 200, loss: 0.1975959986448288
step: 210, loss: 0.03313763812184334
/usr/local/lib/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 4: dev_f1=0.7343750000000001, f1=0.7411764705882354,
best_f1=0.7411764705882354
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05456977337598801
step: 10, loss: 0.18154339492321014
step: 20, loss: 0.03571110963821411
step: 30, loss: 0.10066529363393784
step: 40, loss: 0.1407400667667389
step: 50, loss: 0.06739453971385956
step: 60, loss: 0.05079277232289314
step: 70, loss: 0.21350084245204926
step: 80, loss: 0.018980631604790688
step: 90, loss: 0.04342062398791313
step: 100, loss: 0.05694541335105896
step: 110, loss: 0.012044582515954971
step: 120, loss: 0.05697346851229668
step: 130, loss: 0.03663432225584984
step: 140, loss: 0.04303383454680443

```

```

step: 150, loss: 0.08479883521795273
step: 160, loss: 0.014536177739501
step: 170, loss: 0.0371665433049202
step: 180, loss: 0.09154484421014786
step: 190, loss: 0.05104678496718407
step: 200, loss: 0.07634523510932922
step: 210, loss: 0.019632387906312943
/usr/local/lib/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=0.7366412213740458, f1=0.7649253731343284,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05132793262600899
step: 10, loss: 0.03960135579109192
step: 20, loss: 0.011272263713181019
step: 30, loss: 0.10111445188522339
step: 40, loss: 0.00902779120951891
step: 50, loss: 0.012589870020747185
step: 60, loss: 0.13743463158607483
step: 70, loss: 0.004295657854527235
step: 80, loss: 0.06892911344766617
step: 90, loss: 0.210024893283844
step: 100, loss: 0.03457876294851303
step: 110, loss: 0.04275937005877495
step: 120, loss: 0.007883678190410137
step: 130, loss: 0.04063960164785385
step: 140, loss: 0.0676305890083313
step: 150, loss: 0.010246304795145988
step: 160, loss: 0.08998207747936249
step: 170, loss: 0.07814836502075195
step: 180, loss: 0.048237890005111694
step: 190, loss: 0.17518672347068787
step: 200, loss: 0.010371631011366844
step: 210, loss: 0.032011404633522034
/usr/local/lib/python3.7/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=0.7269155206286837, f1=0.7368421052631579,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02734481170773506
step: 10, loss: 0.08432763069868088
step: 20, loss: 0.03576485440135002
step: 30, loss: 0.053865160793066025
step: 40, loss: 0.03215659409761429
step: 50, loss: 0.07194898277521133
step: 60, loss: 0.15195591747760773
step: 70, loss: 0.006233219522982836
step: 80, loss: 0.07478222995996475
step: 90, loss: 0.00556965172290802
step: 100, loss: 0.02434477023780346
step: 110, loss: 0.04977290332317352
step: 120, loss: 0.171061709523201
step: 130, loss: 0.007219908758997917
step: 140, loss: 0.0013753743842244148
step: 150, loss: 0.04978285729885101
step: 160, loss: 0.021569184958934784
step: 170, loss: 0.023967402055859566
step: 180, loss: 0.00587058812379837
step: 190, loss: 0.01499291229993105
step: 200, loss: 0.1122390553355217
step: 210, loss: 0.061527278274297714
/usr/local/lib/python3.7/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=0.7043795620437955, f1=0.7340425531914894,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029043717309832573
step: 10, loss: 0.021286610513925552
step: 20, loss: 0.03289984539151192
step: 30, loss: 0.02204098179936409
step: 40, loss: 0.11159766465425491
step: 50, loss: 0.00419617909938097
step: 60, loss: 0.20434248447418213
step: 70, loss: 0.004398665390908718
step: 80, loss: 0.008962076157331467
step: 90, loss: 0.017399068921804428
step: 100, loss: 0.007881956174969673
step: 110, loss: 0.02216356061398983
step: 120, loss: 0.009360835887491703
step: 130, loss: 0.01649782620370388
step: 140, loss: 0.011235227808356285
step: 150, loss: 0.06919905543327332
step: 160, loss: 0.05420961230993271

```

```

step: 170, loss: 0.05405333638191223
step: 180, loss: 0.015824411064386368
step: 190, loss: 0.07772200554609299
step: 200, loss: 0.15202659368515015
step: 210, loss: 0.10682608187198639
/usr/local/lib/python3.7/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=0.7037773359840954, f1=0.7311411992263055,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012268549762666225
step: 10, loss: 0.08895570039749146
step: 20, loss: 0.05702917277812958
step: 30, loss: 0.018279781565070152
step: 40, loss: 0.10792545974254608
step: 50, loss: 0.024612177163362503
step: 60, loss: 0.06976667791604996
step: 70, loss: 0.0504579171538353
step: 80, loss: 0.03499895706772804
step: 90, loss: 0.054145459085702896
step: 100, loss: 0.0537886880338192
step: 110, loss: 0.023660913109779358
step: 120, loss: 0.004220946226269007
step: 130, loss: 0.00833418220281601
step: 140, loss: 0.009630898013710976
step: 150, loss: 0.03449680656194687
step: 160, loss: 0.0038066343404352665
step: 170, loss: 0.24866829812526703
step: 180, loss: 0.05688285827636719
step: 190, loss: 0.07304598391056061
step: 200, loss: 0.05266500636935234
step: 210, loss: 0.18463757634162903
/usr/local/lib/python3.7/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=0.7058823529411765, f1=0.7298747763864043,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026095649227499962
step: 10, loss: 0.024552948772907257
step: 20, loss: 0.06951351463794708
step: 30, loss: 0.0115605928003788
step: 40, loss: 0.06741662323474884

```



```

step: 50, loss: 0.08423524349927902
step: 60, loss: 0.002617291174829006
step: 70, loss: 0.08229503035545349
step: 80, loss: 0.020641159266233444
step: 90, loss: 0.0905086025595665
step: 100, loss: 0.0010878859320655465
step: 110, loss: 0.009934778325259686
step: 120, loss: 0.007521617226302624
step: 130, loss: 0.001289478619582951
step: 140, loss: 0.040994614362716675
step: 150, loss: 0.017877932637929916
step: 160, loss: 0.106845423579216
step: 170, loss: 0.0839390754699707
step: 180, loss: 0.007108632940798998
step: 190, loss: 0.19091123342514038
step: 200, loss: 0.03367943689227104
step: 210, loss: 0.011113976128399372
/usr/local/lib/python3.7/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=0.7037037037037037, f1=0.7269372693726938,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009339283220469952
step: 10, loss: 0.05244603008031845
step: 20, loss: 0.0038928533904254436
step: 30, loss: 0.040003325790166855
step: 40, loss: 0.06611429154872894
step: 50, loss: 0.00869066547602415
step: 60, loss: 0.03026590868830681
step: 70, loss: 0.003359393449500203
step: 80, loss: 0.18628832697868347
step: 90, loss: 0.19076456129550934
step: 100, loss: 0.010234195739030838
step: 110, loss: 0.00844673439860344
step: 120, loss: 0.001599954441189766
step: 130, loss: 0.039786018431186676
step: 140, loss: 0.0098541509360075
step: 150, loss: 0.04446427896618843
step: 160, loss: 0.02196681685745716
step: 170, loss: 0.1255486011505127
step: 180, loss: 0.05344081297516823
step: 190, loss: 0.005947571247816086
step: 200, loss: 0.00045198609586805105
step: 210, loss: 0.04915319010615349
/usr/local/lib/python3.7/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=0.708029197080292, f1=0.7298747763864043,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013056501047685742
step: 10, loss: 0.010286500677466393
step: 20, loss: 0.022992562502622604
step: 30, loss: 0.0014142959844321012
step: 40, loss: 0.07307599484920502
step: 50, loss: 0.0459505133330822
step: 60, loss: 0.05803409218788147
step: 70, loss: 0.02360825426876545
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.009269320406019688
step: 90, loss: 0.011512295342981815
step: 100, loss: 0.0029870725702494383
step: 110, loss: 0.0037111560814082623
step: 120, loss: 0.19948714971542358
step: 130, loss: 0.15223723649978638
step: 140, loss: 0.0018031469080597162
step: 150, loss: 0.004450762178748846
step: 160, loss: 0.004157746210694313
step: 170, loss: 0.006952758878469467
step: 180, loss: 0.0018161531770601869
step: 190, loss: 0.03574616461992264
step: 200, loss: 0.04473097249865532
step: 210, loss: 0.004399926867336035
/usr/local/lib/python3.7/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=0.7090558766859344, f1=0.7283236994219653,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01041248720139265
step: 10, loss: 0.004360988736152649
step: 20, loss: 0.0762733519077301
step: 30, loss: 0.058228328824043274
step: 40, loss: 0.020766571164131165
step: 50, loss: 0.0812162235379219
step: 60, loss: 0.02170330286026001
step: 70, loss: 0.16089807450771332
step: 80, loss: 0.0060461354441940784
step: 90, loss: 0.018107786774635315

```

```

step: 100, loss: 0.13944409787654877
step: 110, loss: 0.0008805730612948537
step: 120, loss: 0.02019581012427807
step: 130, loss: 0.002824780298396945
step: 140, loss: 0.019310925155878067
step: 150, loss: 0.0008692471892572939
step: 160, loss: 0.04775167256593704
step: 170, loss: 0.04439841955900192
step: 180, loss: 0.06194370985031128
step: 190, loss: 0.0039033147040754557
step: 200, loss: 0.005758263636380434
step: 210, loss: 0.012087023817002773
/usr/local/lib/python3.7/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=0.7054108216432866, f1=0.733606557377049,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014396612532436848
step: 10, loss: 0.006470724008977413
step: 20, loss: 0.00445956876501441
step: 30, loss: 0.16311345994472504
step: 40, loss: 0.010873008519411087
step: 50, loss: 0.012963675893843174
step: 60, loss: 0.020614320412278175
step: 70, loss: 0.01763913966715336
step: 80, loss: 0.10500291734933853
step: 90, loss: 0.03968773037195206
step: 100, loss: 0.002223716350272298
step: 110, loss: 0.014526656828820705
step: 120, loss: 0.011199595406651497
step: 130, loss: 0.0006457916460931301
step: 140, loss: 0.0009455895633436739
step: 150, loss: 0.05441075935959816
step: 160, loss: 0.0018508005887269974
step: 170, loss: 0.09767590463161469
step: 180, loss: 0.0023871345911175013
step: 190, loss: 0.010574689134955406
step: 200, loss: 0.0006312446785159409
step: 210, loss: 0.025494402274489403
/usr/local/lib/python3.7/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=0.7073608617594256, f1=0.721830985915493,
best_f1=0.7649253731343284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021815301850438118
step: 10, loss: 0.004177126567810774
step: 20, loss: 0.0015388437313959002
step: 30, loss: 0.0004515036125667393
step: 40, loss: 0.0009345709695480764
step: 50, loss: 0.051456790417432785
step: 60, loss: 0.24222376942634583
step: 70, loss: 0.003232471412047744
step: 80, loss: 0.001347905956208706
step: 90, loss: 0.0024477997794747353
step: 100, loss: 0.017095046117901802
step: 110, loss: 0.0019207425648346543
step: 120, loss: 0.02792206220328808
step: 130, loss: 0.018617138266563416
step: 140, loss: 0.002741126576438546
step: 150, loss: 0.0036111718509346247
step: 160, loss: 0.04682154953479767
step: 170, loss: 0.047124799340963364
step: 180, loss: 0.014119233936071396
step: 190, loss: 0.09378570318222046
step: 200, loss: 0.018167143687605858
step: 210, loss: 0.022172311320900917
/usr/local/lib/python3.7/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=0.7129798903107862, f1=0.7253141831238779,
best_f1=0.7649253731343284
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, 351.27it/s]
load_f1 = 0.724907063197026
real_f1 = 0.7238805970149252
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, 264.70it/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: 0.8627178072929382
```

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

```
step: 10, loss: 0.16671040654182434
```

```
step: 20, loss: 0.14568842947483063
```

```
step: 30, loss: 0.48883506655693054
```

```

step: 40, loss: 0.24801698327064514
step: 50, loss: 0.3008936941623688
step: 60, loss: 0.36271533370018005
step: 70, loss: 0.1778227835893631
step: 80, loss: 0.4913623332977295
step: 90, loss: 0.23029959201812744
step: 100, loss: 0.21411846578121185
step: 110, loss: 0.2284257709980011
step: 120, loss: 0.40560638904571533
step: 130, loss: 0.36715438961982727
step: 140, loss: 0.31080660223960876
step: 150, loss: 0.23723262548446655
step: 160, loss: 0.2174728661775589
step: 170, loss: 0.38659611344337463
step: 180, loss: 0.1954127848148346
step: 190, loss: 0.12269803881645203
/usr/local/lib/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=0.6358695652173914, f1=0.6666666666666667,
best_f1=0.6666666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1846609264612198
step: 10, loss: 0.041841909289360046
step: 20, loss: 0.03439252823591232
step: 30, loss: 0.06597258895635605
step: 40, loss: 0.3439084589481354
step: 50, loss: 0.2657395899295807
step: 60, loss: 0.16082477569580078
step: 70, loss: 0.22780027985572815
step: 80, loss: 0.11884629726409912
step: 90, loss: 0.1314549446105957
step: 100, loss: 0.22057124972343445
step: 110, loss: 0.07661408185958862
step: 120, loss: 0.10674259066581726
step: 130, loss: 0.14211644232273102
step: 140, loss: 0.11090680211782455
step: 150, loss: 0.1456596553325653
step: 160, loss: 0.012904475443065166
step: 170, loss: 0.0677865520119667
step: 180, loss: 0.10434424132108688
step: 190, loss: 0.09172561764717102

```

```

/usr/local/lib/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=0.7830687830687831, f1=0.8031914893617021,
best_f1=0.8031914893617021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0705532506108284
step: 10, loss: 0.1934116780757904
step: 20, loss: 0.09445696324110031
step: 30, loss: 0.02896285429596901
step: 40, loss: 0.04100850224494934
step: 50, loss: 0.26546069979667664
step: 60, loss: 0.018902959302067757
step: 70, loss: 0.11868923902511597
step: 80, loss: 0.123543381690979
step: 90, loss: 0.0264650359749794
step: 100, loss: 0.06640613824129105
step: 110, loss: 0.20141376554965973
step: 120, loss: 0.019741088151931763
step: 130, loss: 0.008932274766266346
step: 140, loss: 0.01199459657073021
step: 150, loss: 0.10101401060819626
step: 160, loss: 0.20857571065425873
step: 170, loss: 0.021583108231425285
step: 180, loss: 0.053742583841085434
step: 190, loss: 0.14046689867973328
/usr/local/lib/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=0.8021978021978022, f1=0.753623188405797,
best_f1=0.753623188405797
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.052109770476818085
step: 10, loss: 0.05480457842350006
step: 20, loss: 0.025461547076702118
step: 30, loss: 0.021238358691334724

```



```

step: 40, loss: 0.01278295461088419
step: 50, loss: 0.023260045796632767
step: 60, loss: 0.11680007725954056
step: 70, loss: 0.015450721606612206
step: 80, loss: 0.085272416472435
step: 90, loss: 0.04034166783094406
step: 100, loss: 0.020392583683133125
step: 110, loss: 0.02529052086174488
step: 120, loss: 0.0774620920419693
step: 130, loss: 0.21326375007629395
step: 140, loss: 0.07497386634349823
step: 150, loss: 0.06469148397445679
step: 160, loss: 0.014637090265750885
step: 170, loss: 0.006553563289344311
step: 180, loss: 0.20556046068668365
step: 190, loss: 0.022990116849541664
/usr/local/lib/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=0.8172588832487311, f1=0.781491002570694,
best_f1=0.781491002570694
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12307664006948471
step: 10, loss: 0.004525531083345413
step: 20, loss: 0.08579140901565552
step: 30, loss: 0.004025572445243597
step: 40, loss: 0.05801628157496452
step: 50, loss: 0.01016329787671566
step: 60, loss: 0.01835128851234913
step: 70, loss: 0.01355783548206091
step: 80, loss: 0.007733989506959915
step: 90, loss: 0.05952227860689163
step: 100, loss: 0.17980660498142242
step: 110, loss: 0.010502181947231293
step: 120, loss: 0.002455555135384202
step: 130, loss: 0.010357226245105267
step: 140, loss: 0.0021948153153061867
step: 150, loss: 0.03903847932815552
step: 160, loss: 0.07129540294408798
step: 170, loss: 0.01940098963677883
step: 180, loss: 0.14768999814987183
step: 190, loss: 0.15656881034374237

```

```

/usr/local/lib/python3.7/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=0.8042895442359249, f1=0.7967032967032966,
best_f1=0.781491002570694
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13079029321670532
step: 10, loss: 0.0037693886552006006
step: 20, loss: 0.020893096923828125
step: 30, loss: 0.043324995785951614
step: 40, loss: 0.026863664388656616
step: 50, loss: 0.021747678518295288
step: 60, loss: 0.0013024128274992108
step: 70, loss: 0.011168483644723892
step: 80, loss: 0.04935229942202568
step: 90, loss: 0.023217782378196716
step: 100, loss: 0.0017665078630670905
step: 110, loss: 0.10449021309614182
step: 120, loss: 0.0036311016883701086
step: 130, loss: 0.008703736588358879
step: 140, loss: 0.0012715638149529696
step: 150, loss: 0.008566995151340961
step: 160, loss: 0.004081660881638527
step: 170, loss: 0.03451430797576904
step: 180, loss: 0.00925035122781992
step: 190, loss: 0.04057640954852104
/usr/local/lib/python3.7/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=0.8116710875331565, f1=0.7901907356948229,
best_f1=0.781491002570694
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006313602440059185
step: 10, loss: 0.015498004853725433
step: 20, loss: 0.0008380282670259476
step: 30, loss: 0.002426300197839737
step: 40, loss: 0.01294996589422226
step: 50, loss: 0.01982567273080349
step: 60, loss: 0.026224473491311073
step: 70, loss: 0.020414885133504868
step: 80, loss: 0.005829979665577412
step: 90, loss: 0.13477100431919098
step: 100, loss: 0.0025087937247008085
step: 110, loss: 0.030672797933220863

```

```

step: 120, loss: 0.003446130082011223
step: 130, loss: 0.005599323660135269
step: 140, loss: 0.004452477674931288
step: 150, loss: 0.0069260080344974995
step: 160, loss: 0.0006048373761586845
step: 170, loss: 0.019877120852470398
step: 180, loss: 0.00399048812687397
step: 190, loss: 0.015349432826042175
/usr/local/lib/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=0.8203753351206434, f1=0.7726027397260273,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03119955025613308
step: 10, loss: 0.011561179533600807
step: 20, loss: 0.003932773135602474
step: 30, loss: 0.07945328950881958
step: 40, loss: 0.010546070523560047
step: 50, loss: 0.0017979337135329843
step: 60, loss: 0.0013756523840129375
step: 70, loss: 0.004935505334287882
step: 80, loss: 0.1671147346496582
step: 90, loss: 0.0039251623675227165
step: 100, loss: 0.05119534209370613
step: 110, loss: 0.004882526118308306
step: 120, loss: 0.0010341345332562923
step: 130, loss: 0.0035176605451852083
step: 140, loss: 0.0013085745740681887
step: 150, loss: 0.002883962821215391
step: 160, loss: 0.0026489917654544115
step: 170, loss: 0.0027235914021730423
step: 180, loss: 0.051134683191776276
step: 190, loss: 0.030752716585993767
/usr/local/lib/python3.7/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=0.8064516129032259, f1=0.8206521739130436,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.0032365082297474146
step: 10, loss: 0.009887589141726494
step: 20, loss: 0.19259677827358246
step: 30, loss: 0.0025320251006633043
step: 40, loss: 0.0011311317794024944
step: 50, loss: 0.0005684599746018648
step: 60, loss: 0.0018060934962704778
step: 70, loss: 0.00291437073610723
step: 80, loss: 0.14400406181812286
step: 90, loss: 0.0732576996088028
step: 100, loss: 0.007006341591477394
step: 110, loss: 0.0052689253352582455
step: 120, loss: 0.0745810940861702
step: 130, loss: 0.004005130846053362
step: 140, loss: 0.0015431003412231803
step: 150, loss: 0.003037230344489217
step: 160, loss: 0.00059013586724177
step: 170, loss: 0.0855337530374527
step: 180, loss: 0.00866785366088152
step: 190, loss: 0.027152959257364273
/usr/local/lib/python3.7/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=0.7969924812030076, f1=0.7939698492462313,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007256606477312744
step: 10, loss: 0.024889053776860237
step: 20, loss: 0.0023562107235193253
step: 30, loss: 0.0012013863306492567
step: 40, loss: 0.0005004349513910711
step: 50, loss: 0.0008014248451218009
step: 60, loss: 0.008888537995517254
step: 70, loss: 0.020777998492121696
step: 80, loss: 0.05072972550988197
step: 90, loss: 0.0008708004024811089
step: 100, loss: 0.01277873758226633
step: 110, loss: 0.0130890728905797
step: 120, loss: 0.002214850392192602
step: 130, loss: 0.0033088193740695715
step: 140, loss: 0.0008061190019361675
step: 150, loss: 0.00033621597685851157
step: 160, loss: 0.01765129528939724
step: 170, loss: 0.02466779761016369
step: 180, loss: 0.0023166988976299763
step: 190, loss: 0.003164220368489623

```

```

/usr/local/lib/python3.7/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=0.79155672823219, f1=0.7771739130434783,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000570993812289089
step: 10, loss: 0.021757075563073158
step: 20, loss: 0.008985916152596474
step: 30, loss: 0.027351433411240578
step: 40, loss: 0.009731951169669628
step: 50, loss: 0.004597735125571489
step: 60, loss: 0.0010108056012541056
step: 70, loss: 0.004021424800157547
step: 80, loss: 0.002142448676750064
step: 90, loss: 0.025614479556679726
step: 100, loss: 0.00040354736847802997
step: 110, loss: 0.002425726503133774
step: 120, loss: 0.006511061452329159
step: 130, loss: 0.0004276452527847141
step: 140, loss: 0.0005099060945212841
step: 150, loss: 0.04046812653541565
step: 160, loss: 0.003922808449715376
step: 170, loss: 0.001084098475985229
step: 180, loss: 0.12271533906459808
step: 190, loss: 0.002470723120495677
/usr/local/lib/python3.7/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=0.7990074441687345, f1=0.7839195979899498,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012818466639146209
step: 10, loss: 0.0022889275569468737
step: 20, loss: 0.011486247181892395
step: 30, loss: 0.002239733701571822
step: 40, loss: 0.11822743713855743
step: 50, loss: 0.04992276802659035
step: 60, loss: 0.0010678820544853806
step: 70, loss: 0.0010451850248500705
step: 80, loss: 0.01107343751937151
step: 90, loss: 0.0008967944304458797
step: 100, loss: 0.0005786665715277195
step: 110, loss: 0.0008652795222587883

```

```

step: 120, loss: 0.0006097008008509874
step: 130, loss: 0.0003586619859561324
step: 140, loss: 0.0018646438838914037
step: 150, loss: 0.002155229914933443
step: 160, loss: 0.24797165393829346
step: 170, loss: 0.002989706117659807
step: 180, loss: 0.000574654433876276
step: 190, loss: 0.01879388839006424
/usr/local/lib/python3.7/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=0.7958115183246073, f1=0.7826086956521738,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005648134974762797
step: 10, loss: 0.038596659898757935
step: 20, loss: 0.004369268659502268
step: 30, loss: 0.0005990284844301641
step: 40, loss: 0.0005612173117697239
step: 50, loss: 0.0005851213354617357
step: 60, loss: 0.004185175057500601
step: 70, loss: 0.0035219606943428516
step: 80, loss: 0.004359554033726454
step: 90, loss: 0.00200646067969501
step: 100, loss: 0.001166836591437459
step: 110, loss: 0.05529589578509331
step: 120, loss: 0.0026101558469235897
step: 130, loss: 0.011667311191558838
step: 140, loss: 0.0040184082463383675
step: 150, loss: 0.003280816599726677
step: 160, loss: 0.0010583421681076288
step: 170, loss: 0.0007201146217994392
step: 180, loss: 0.009002884849905968
step: 190, loss: 0.010927978903055191
/usr/local/lib/python3.7/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=0.8060453400503779, f1=0.8031496062992125,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024456672836095095
step: 10, loss: 0.0005337994080036879
step: 20, loss: 0.002215284388512373
step: 30, loss: 0.0015995479188859463

```

```

step: 40, loss: 0.001245575025677681
step: 50, loss: 0.003907529171556234
step: 60, loss: 0.0006441805162467062
step: 70, loss: 0.0035926734562963247
step: 80, loss: 0.00042009344906546175
step: 90, loss: 0.0033344682306051254
step: 100, loss: 0.01857158914208412
step: 110, loss: 0.005381532944738865
step: 120, loss: 0.0029030328150838614
step: 130, loss: 0.0006981455953791738
step: 140, loss: 0.0003421749279368669
step: 150, loss: 0.0008572874940000474
step: 160, loss: 0.000769011618103832
step: 170, loss: 0.0030012433417141438
step: 180, loss: 0.0026700578164309263
step: 190, loss: 0.00018283991084899753
/usr/local/lib/python3.7/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=0.8054054054054054, f1=0.7878787878787877,
best_f1=0.7726027397260273
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001077670487575233
step: 10, loss: 0.000513887673150748
step: 20, loss: 0.0011337930336594582
step: 30, loss: 0.00040310181793756783
step: 40, loss: 0.0003035617701243609
step: 50, loss: 0.002005981747061014
step: 60, loss: 0.09511175006628036
step: 70, loss: 0.0020435925107449293
step: 80, loss: 0.00029408864793367684
step: 90, loss: 0.0018404144793748856
step: 100, loss: 0.0004590573953464627
step: 110, loss: 0.0009866802720353007
step: 120, loss: 0.0032170703634619713
step: 130, loss: 0.0004169044259469956
step: 140, loss: 0.0033378060907125473
step: 150, loss: 0.013124535791575909
step: 160, loss: 0.024534469470381737
step: 170, loss: 0.0005026660510338843
step: 180, loss: 0.0005894247442483902
step: 190, loss: 0.01598329097032547
/usr/local/lib/python3.7/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=0.8044077134986225, f1=0.7796610169491526,

```

```

best_f1=0.7726027397260273
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.")
2049it [00:08, 234.64it/s]
load_f1 = 0.7248677248677249
real_f1 = 0.7157360406091372
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, 265.13it/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 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: 0.8550364375114441
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.23063340783119202
step: 20, loss: 0.15151835978031158
step: 30, loss: 0.2318413406610489
step: 40, loss: 0.31542158126831055
step: 50, loss: 0.37664520740509033
step: 60, loss: 0.4431557357311249
step: 70, loss: 0.3083363175392151
step: 80, loss: 0.24777965247631073
step: 90, loss: 0.39317551255226135
step: 100, loss: 0.22290587425231934
step: 110, loss: 0.16165949404239655
step: 120, loss: 0.5010371208190918
step: 130, loss: 0.3293240964412689
step: 140, loss: 0.3148387372493744
step: 150, loss: 0.1809147149324417
step: 160, loss: 0.17412994801998138
step: 170, loss: 0.26381292939186096
/usr/local/lib/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=0.7029478458049887, f1=0.6917647058823531,
best_f1=0.6917647058823531
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.44375714659690857
step: 10, loss: 0.045747071504592896
step: 20, loss: 0.18110762536525726
step: 30, loss: 0.1353013664484024
step: 40, loss: 0.14779649674892426
step: 50, loss: 0.10263186693191528
step: 60, loss: 0.042186591774225235
step: 70, loss: 0.11801076680421829
step: 80, loss: 0.23806293308734894
step: 90, loss: 0.14579644799232483
step: 100, loss: 0.08324705064296722
step: 110, loss: 0.14819872379302979
step: 120, loss: 0.039213016629219055
step: 130, loss: 0.09773208945989609
step: 140, loss: 0.10559969395399094
step: 150, loss: 0.09399275481700897
step: 160, loss: 0.05661141127347946
step: 170, loss: 0.23842228949069977
/usr/local/lib/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=0.7971360381861575, f1=0.7972665148063781,
best_f1=0.7972665148063781
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04655541479587555
step: 10, loss: 0.01739642769098282
step: 20, loss: 0.07260570675134659
step: 30, loss: 0.07913511991500854
step: 40, loss: 0.0028241747058928013
step: 50, loss: 0.1586935669183731
step: 60, loss: 0.15515843033790588
step: 70, loss: 0.052299849689006805
step: 80, loss: 0.09281107783317566
step: 90, loss: 0.06029356271028519
step: 100, loss: 0.05283915251493454
step: 110, loss: 0.011825119145214558
step: 120, loss: 0.0042530233040452
step: 130, loss: 0.09157010167837143
step: 140, loss: 0.003079968737438321
step: 150, loss: 0.02183718979358673
step: 160, loss: 0.010985927656292915
step: 170, loss: 0.18017058074474335

```

```

/usr/local/lib/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=0.8086124401913874, f1=0.7963800904977376,
best_f1=0.7963800904977376
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0739104375243187
step: 10, loss: 0.17698509991168976
step: 20, loss: 0.01385527290403843
step: 30, loss: 0.014809327200055122
step: 40, loss: 0.004548123572021723
step: 50, loss: 0.03174415975809097
step: 60, loss: 0.06651893258094788
step: 70, loss: 0.09201856702566147
step: 80, loss: 0.010432484559714794
step: 90, loss: 0.01043434627354145
step: 100, loss: 0.013091707602143288
step: 110, loss: 0.010748504661023617
step: 120, loss: 0.1902455985546112
step: 130, loss: 0.015837695449590683
step: 140, loss: 0.05027971416711807
step: 150, loss: 0.033939946442842484
step: 160, loss: 0.02492610737681389
step: 170, loss: 0.006075110752135515
/usr/local/lib/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=0.8393782383419689, f1=0.8478802992518701,
best_f1=0.8478802992518701
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020993897691369057
step: 10, loss: 0.010901975445449352
step: 20, loss: 0.002174871740862727
step: 30, loss: 0.00878628809005022
step: 40, loss: 0.13134582340717316
step: 50, loss: 0.0038261900190263987

```

```

step: 60, loss: 0.0008389729191549122
step: 70, loss: 0.018284713849425316
step: 80, loss: 0.009168717078864574
step: 90, loss: 0.033701974898576736
step: 100, loss: 0.020350072532892227
step: 110, loss: 0.0863315612077713
step: 120, loss: 0.015888456255197525
step: 130, loss: 0.033358700573444366
step: 140, loss: 0.026842758059501648
step: 150, loss: 0.02846558392047882
step: 160, loss: 0.002551443874835968
step: 170, loss: 0.025161944329738617
/usr/local/lib/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=0.839907192575406, f1=0.8272727272727274,
best_f1=0.8272727272727274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00873956736177206
step: 10, loss: 0.0536331944167614
step: 20, loss: 0.048956289887428284
step: 30, loss: 0.02577819488942623
step: 40, loss: 0.01795944571495056
step: 50, loss: 0.0031680255196988583
step: 60, loss: 0.06359057873487473
step: 70, loss: 0.0032486203126609325
step: 80, loss: 0.027849366888403893
step: 90, loss: 0.023715175688266754
step: 100, loss: 0.0185635257512331
step: 110, loss: 0.012417410500347614
step: 120, loss: 0.0232840683311224
step: 130, loss: 0.21272839605808258
step: 140, loss: 0.004183050710707903
step: 150, loss: 0.12404340505599976
step: 160, loss: 0.012585926800966263
step: 170, loss: 0.008763485588133335
/usr/local/lib/python3.7/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=0.8108108108108109, f1=0.8426395939086295,
best_f1=0.8272727272727274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0065291933715343475
step: 10, loss: 0.0007270717760547996
step: 20, loss: 0.005688199773430824
step: 30, loss: 0.002045243978500366
step: 40, loss: 0.007284608203917742
step: 50, loss: 0.0024264066014438868
step: 60, loss: 0.14095239341259003
step: 70, loss: 0.0013905189698562026
step: 80, loss: 0.002353043295443058
step: 90, loss: 0.005292234010994434
step: 100, loss: 0.0031213401816785336
step: 110, loss: 0.0009308913722634315
step: 120, loss: 0.2684677243232727
step: 130, loss: 0.08173500001430511
step: 140, loss: 0.04215359315276146
step: 150, loss: 0.025460708886384964
step: 160, loss: 0.008100584149360657
step: 170, loss: 0.10121723264455795
/usr/local/lib/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=0.8407310704960835, f1=0.8459657701711493,
best_f1=0.8459657701711493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08888686448335648
step: 10, loss: 0.004686261527240276
step: 20, loss: 0.0014174559619277716
step: 30, loss: 0.00374490930698812
step: 40, loss: 0.019440501928329468
step: 50, loss: 0.00035742708132602274
step: 60, loss: 0.002498813671991229
step: 70, loss: 0.0734039843082428
step: 80, loss: 0.00030992948450148106
step: 90, loss: 0.1047942042350769
step: 100, loss: 0.0031172994058579206
step: 110, loss: 0.029815377667546272
step: 120, loss: 0.003535136580467224
step: 130, loss: 0.004016845487058163
step: 140, loss: 0.07666922360658646
step: 150, loss: 0.07068704068660736

```

```

step: 160, loss: 0.028597692027688026
step: 170, loss: 0.005539513658732176
/usr/local/lib/python3.7/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=0.8373333333333333, f1=0.8514851485148516,
best_f1=0.8459657701711493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009236459620296955
step: 10, loss: 0.1633760631084442
step: 20, loss: 0.0005970189231447875
step: 30, loss: 0.009328319691121578
step: 40, loss: 0.028775887563824654
step: 50, loss: 0.012377961538732052
step: 60, loss: 0.000875726982485503
step: 70, loss: 0.025887377560138702
step: 80, loss: 0.08749502152204514
step: 90, loss: 0.002336375415325165
step: 100, loss: 0.026567528024315834
step: 110, loss: 0.0021793560590595007
step: 120, loss: 0.03386671468615532
step: 130, loss: 0.00014906386786606163
step: 140, loss: 0.0013999615330249071
step: 150, loss: 0.013197099789977074
step: 160, loss: 0.011234739795327187
step: 170, loss: 0.03907477483153343
/usr/local/lib/python3.7/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=0.8320413436692506, f1=0.8536585365853658,
best_f1=0.8459657701711493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003263051330577582
step: 10, loss: 0.03196696937084198
step: 20, loss: 0.00014046703290659934
step: 30, loss: 0.00034517896710895
step: 40, loss: 0.0052107227966189384
step: 50, loss: 0.020809590816497803
step: 60, loss: 0.0008105384767986834
step: 70, loss: 0.0011841424275189638
step: 80, loss: 0.0041723777540028095
step: 90, loss: 0.025226334109902382
step: 100, loss: 0.00045826053246855736
step: 110, loss: 0.0014200832229107618

```

```

step: 120, loss: 0.037520963698625565
step: 130, loss: 0.005783423315733671
step: 140, loss: 0.008853638544678688
step: 150, loss: 0.03481225669384003
step: 160, loss: 0.0007335599511861801
step: 170, loss: 0.00086647440912202
/usr/local/lib/python3.7/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=0.8381962864721485, f1=0.8459657701711493,
best_f1=0.8459657701711493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00024254026357084513
step: 10, loss: 0.005936273839324713
step: 20, loss: 0.012471749447286129
step: 30, loss: 0.018812837079167366
step: 40, loss: 0.009377784095704556
step: 50, loss: 0.0013911445857957006
step: 60, loss: 0.0006041182787157595
step: 70, loss: 0.0040815346874296665
step: 80, loss: 0.00018949910008814186
step: 90, loss: 0.002792813815176487
step: 100, loss: 0.0001545955310575664
step: 110, loss: 0.0020359582267701626
step: 120, loss: 0.006308858282864094
step: 130, loss: 0.0014026087010279298
step: 140, loss: 0.02776339277625084
step: 150, loss: 0.0031449191737920046
step: 160, loss: 0.0008554124506190419
step: 170, loss: 0.0461617149412632
/usr/local/lib/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=0.8428927680798006, f1=0.8408551068883611,
best_f1=0.8408551068883611
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036464256700128317
step: 10, loss: 0.004057563375681639
step: 20, loss: 0.026115233078598976
step: 30, loss: 0.00272091431543231

```



```

step: 40, loss: 8.005578274605796e-05
step: 50, loss: 0.014150544069707394
step: 60, loss: 0.00016607709403615445
step: 70, loss: 0.012261058203876019
step: 80, loss: 0.10387516021728516
step: 90, loss: 0.00023227774363476783
step: 100, loss: 0.00343851069919765
step: 110, loss: 0.0001228336477652192
step: 120, loss: 0.005037053022533655
step: 130, loss: 0.0012977279257029295
step: 140, loss: 0.00022948150581214577
step: 150, loss: 0.0005203927285037935
step: 160, loss: 0.011956078931689262
step: 170, loss: 0.0012730623129755259
/usr/local/lib/python3.7/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=0.8383838383838383, f1=0.8448687350835321,
best_f1=0.8408551068883611
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037950166733935475
step: 10, loss: 0.07981758564710617
step: 20, loss: 0.00444795424118638
step: 30, loss: 0.018495261669158936
step: 40, loss: 6.972498522372916e-05
step: 50, loss: 0.014840522781014442
step: 60, loss: 0.0003258013166487217
step: 70, loss: 0.014352872036397457
step: 80, loss: 0.0004886166425421834
step: 90, loss: 0.020264366641640663
step: 100, loss: 6.30241702310741e-05
step: 110, loss: 0.0007758245337754488
step: 120, loss: 0.0016706492751836777
step: 130, loss: 0.009742732159793377
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 140, loss: 0.0017707310616970062
step: 150, loss: 0.00012279252405278385
step: 160, loss: 0.010167808271944523
step: 170, loss: 0.0032798710744827986
/usr/local/lib/python3.7/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=0.8400954653937948, f1=0.8222222222222222,
best_f1=0.8408551068883611
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00027167139342054725
step: 10, loss: 0.00017562428547535092
step: 20, loss: 0.10355507582426071
step: 30, loss: 0.001827575615607202
step: 40, loss: 0.001146976719610393
step: 50, loss: 0.0013045946834608912
step: 60, loss: 0.0015370309120044112
step: 70, loss: 0.006518900394439697
step: 80, loss: 0.0015667605912312865
step: 90, loss: 0.0007459412445314229
step: 100, loss: 8.044099377002567e-05
step: 110, loss: 0.0011727178934961557
step: 120, loss: 0.0003042207099497318
step: 130, loss: 0.0008614715188741684
step: 140, loss: 0.0001971068704733625
step: 150, loss: 0.0004760223673656583
step: 160, loss: 0.0214955136179924
step: 170, loss: 0.00018258279305882752
/usr/local/lib/python3.7/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=0.8396946564885497, f1=0.8523002421307506,
best_f1=0.8408551068883611
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0026320968754589558
step: 10, loss: 0.010307841002941132
step: 20, loss: 0.02082614041864872
step: 30, loss: 0.02536880411207676
step: 40, loss: 7.320585427805781e-05
step: 50, loss: 0.004437961149960756
step: 60, loss: 4.0097922465065494e-05
step: 70, loss: 0.017879357561469078
step: 80, loss: 0.0006175617454573512
step: 90, loss: 8.099646947812289e-05
step: 100, loss: 0.0618363693356514
step: 110, loss: 0.00010018863395089284
step: 120, loss: 0.0040743364952504635
step: 130, loss: 0.00047747965436428785
step: 140, loss: 0.00022531268768943846
step: 150, loss: 0.05341412499547005
step: 160, loss: 5.55986407562159e-05
step: 170, loss: 0.0035147052258253098
/usr/local/lib/python3.7/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=0.8396946564885497, f1=0.8523002421307506,
best_f1=0.8408551068883611
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 [00:05, 330.39it/s]
```

```
load_f1 = 0.30866141732283464
```

```
real_f1 = 0.30913348946135827
```

```
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, 261.23it/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/
```

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        : 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: 0.8231320381164551
step: 10, loss: 0.45563408732414246
step: 20, loss: 0.5584948062896729
step: 30, loss: 0.3338666260242462
step: 40, loss: 0.14534421265125275
step: 50, loss: 0.31031882762908936
step: 60, loss: 0.05943259969353676
step: 70, loss: 0.2205013930797577
step: 80, loss: 0.2699096202850342
step: 90, loss: 0.081609345972538
step: 100, loss: 0.05086769536137581
step: 110, loss: 0.04454552009701729
step: 120, loss: 0.010038565844297409
step: 130, loss: 0.005029789637774229
step: 140, loss: 0.0032395338639616966
step: 150, loss: 0.2240203469991684
step: 160, loss: 0.15010863542556763
step: 170, loss: 0.01210672220778465
step: 180, loss: 0.013675792142748833
step: 190, loss: 0.028492070734500885
step: 200, loss: 0.008172855712473392
step: 210, loss: 0.008182254619896412
step: 220, loss: 0.011813189834356308
step: 230, loss: 0.010451515205204487
```

```

/usr/local/lib/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=0.9735682819383259, f1=0.9712389380530975,
best_f1=0.9712389380530975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007418667431920767
step: 10, loss: 0.0014182173181325197
step: 20, loss: 0.003410050179809332
step: 30, loss: 0.0031887413933873177
step: 40, loss: 0.0034827599301934242
step: 50, loss: 0.059642158448696136
step: 60, loss: 0.0058814953081309795
step: 70, loss: 0.13334128260612488
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.002164010191336274
step: 90, loss: 0.008840978145599365
step: 100, loss: 0.0756957083940506
step: 110, loss: 0.11323271691799164
step: 120, loss: 0.0021326635032892227
step: 130, loss: 0.005706092342734337
step: 140, loss: 0.3180291950702667
step: 150, loss: 0.012495744042098522
step: 160, loss: 0.014712728559970856
step: 170, loss: 0.007893827743828297
step: 180, loss: 0.009200690314173698
step: 190, loss: 0.040443457663059235
step: 200, loss: 0.006528449710458517
step: 210, loss: 0.033470455557107925
step: 220, loss: 0.0019285097951069474
step: 230, loss: 0.01800977624952793
/usr/local/lib/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=0.9810479375696767, f1=0.9754464285714286,
best_f1=0.9754464285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01792755536735058
step: 10, loss: 0.011989359743893147
step: 20, loss: 0.01080658845603466
step: 30, loss: 0.06282275170087814
step: 40, loss: 0.07898560911417007
step: 50, loss: 0.004025045782327652
step: 60, loss: 0.0007377955480478704
step: 70, loss: 0.0025954334996640682
step: 80, loss: 0.0008554158848710358
step: 90, loss: 0.03027297928929329
step: 100, loss: 0.0024189299438148737
step: 110, loss: 0.02384653501212597
step: 120, loss: 0.19886822998523712
step: 130, loss: 0.0006758314557373524
step: 140, loss: 0.003082968294620514
step: 150, loss: 0.0009605936356820166
step: 160, loss: 0.0018255806062370539
step: 170, loss: 0.0064781224355101585
step: 180, loss: 0.005707699805498123
step: 190, loss: 0.0016754524549469352
step: 200, loss: 0.0021929601207375526
step: 210, loss: 0.013241397216916084
step: 220, loss: 0.0039335694164037704
step: 230, loss: 0.09026995301246643
/usr/local/lib/python3.7/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=0.9723756906077348, f1=0.9691629955947136,
best_f1=0.9754464285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002796497428789735
step: 10, loss: 0.00541086308658123
step: 20, loss: 0.0007469747797586024
step: 30, loss: 0.004155739210546017
step: 40, loss: 0.004673525225371122
step: 50, loss: 0.0037820523139089346
step: 60, loss: 0.0037833568640053272
step: 70, loss: 0.0056318724527955055
step: 80, loss: 0.08085296303033829
step: 90, loss: 0.043793804943561554
step: 100, loss: 0.0008316010353155434
step: 110, loss: 0.007314004469662905
step: 120, loss: 0.004337466321885586
step: 130, loss: 0.002023432869464159
step: 140, loss: 0.0007164974231272936

```

```

step: 150, loss: 0.0013034598669037223
step: 160, loss: 0.0022955555468797684
step: 170, loss: 0.00979414489120245
step: 180, loss: 0.09481513500213623
step: 190, loss: 0.015061816200613976
step: 200, loss: 0.001956525258719921
step: 210, loss: 0.00047881784848868847
step: 220, loss: 0.0008255133870989084
step: 230, loss: 0.0026672089006751776
/usr/local/lib/python3.7/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=0.9776286353467561, f1=0.9777777777777777,
best_f1=0.9754464285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01534641906619072
step: 10, loss: 0.0034671796020120382
step: 20, loss: 0.001153492950834334
step: 30, loss: 0.0005635273992083967
step: 40, loss: 0.00042318389751017094
step: 50, loss: 0.00037923300988040864
step: 60, loss: 0.0003607243415899575
step: 70, loss: 0.0022754098754376173
step: 80, loss: 0.00040867767529562116
step: 90, loss: 0.014078781008720398
step: 100, loss: 0.0011307087261229753
step: 110, loss: 0.0013788628857582808
step: 120, loss: 0.023983964696526527
step: 130, loss: 0.05223919078707695
step: 140, loss: 0.0004352530522737652
step: 150, loss: 0.00035378793836571276
step: 160, loss: 0.07843257486820221
step: 170, loss: 0.14865782856941223
step: 180, loss: 0.0028782400768250227
step: 190, loss: 0.0013743859017267823
step: 200, loss: 0.00336071802303195
step: 210, loss: 0.00042170562664978206
step: 220, loss: 0.0005132497753947973
step: 230, loss: 0.0005967881297692657
/usr/local/lib/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=0.9831649831649831, 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: 0.0012235940666869283
step: 10, loss: 0.00039473181823268533
step: 20, loss: 0.0006066488567739725
step: 30, loss: 0.001169196330010891
step: 40, loss: 0.0022242118138819933
step: 50, loss: 0.012039466761052608
step: 60, loss: 0.006190434098243713
step: 70, loss: 0.0003694758052006364
step: 80, loss: 0.00046497449511662126
step: 90, loss: 0.000490016071125865
step: 100, loss: 0.002544933697208762
step: 110, loss: 0.03042548894882202
step: 120, loss: 0.0003791718918364495
step: 130, loss: 0.00022973337036091834
step: 140, loss: 0.0004979536170139909
step: 150, loss: 0.00024020503042265773
step: 160, loss: 0.008869467303156853
step: 170, loss: 0.00041662080911919475
step: 180, loss: 0.0004059710481669754
step: 190, loss: 0.0003783715656027198
step: 200, loss: 0.0021560380700975657
step: 210, loss: 0.0003850353532470763
step: 220, loss: 0.0003051279636565596
step: 230, loss: 0.0047552576288580894
/usr/local/lib/python3.7/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=0.9831271091113611, f1=0.9776286353467561,
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: 0.05457064136862755
step: 10, loss: 0.0008487111772410572
step: 20, loss: 0.00025205040583387017
step: 30, loss: 0.0009579192847013474
step: 40, loss: 0.0015591088449582458
step: 50, loss: 0.0011277315206825733
step: 60, loss: 0.10762762278318405
step: 70, loss: 0.020616548135876656
step: 80, loss: 7.486938557121903e-05
step: 90, loss: 0.00035916612250730395
step: 100, loss: 0.000885224319063127

```

```

step: 110, loss: 0.10644899308681488
step: 120, loss: 0.0038783964700996876
step: 130, loss: 0.0013783330796286464
step: 140, loss: 0.0001734532561386004
step: 150, loss: 0.0011993941152468324
step: 160, loss: 0.00019914934819098562
step: 170, loss: 0.00013270256749819964
step: 180, loss: 0.00027097866404801607
step: 190, loss: 0.00027288158889859915
step: 200, loss: 0.0005653490079566836
step: 210, loss: 0.00010722781007643789
step: 220, loss: 0.0003902660682797432
step: 230, loss: 0.01952873170375824
/usr/local/lib/python3.7/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=0.9730337078651685, f1=0.967452300785634,
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: 0.020684534683823586
step: 10, loss: 0.01689787209033966
step: 20, loss: 0.00022116488253232092
step: 30, loss: 0.00032895526965148747
step: 40, loss: 0.10888688266277313
step: 50, loss: 0.0023221697192639112
step: 60, loss: 0.015085455030202866
step: 70, loss: 0.0002264241484226659
step: 80, loss: 0.0008449435699731112
step: 90, loss: 0.00014255972928367555
step: 100, loss: 0.00016259460244327784
step: 110, loss: 0.00019742776930797845
step: 120, loss: 6.026513801771216e-05
step: 130, loss: 0.0007539728539995849
step: 140, loss: 0.0002086201129714027
step: 150, loss: 0.0024760256055742502
step: 160, loss: 0.0011272374540567398
step: 170, loss: 0.0024978145956993103
step: 180, loss: 0.0070367068983614445
step: 190, loss: 0.0003971756377723068
step: 200, loss: 0.020504586398601532
step: 210, loss: 0.0017480073729529977
step: 220, loss: 0.0002836367057170719
step: 230, loss: 0.03733384981751442
/usr/local/lib/python3.7/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=0.9777777777777777, f1=0.9700996677740864,
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: 0.00024850762565620244
step: 10, loss: 0.000283287517959252
step: 20, loss: 0.0005566218169406056
step: 30, loss: 0.0002850882592611015
step: 40, loss: 0.001656350097618997
step: 50, loss: 0.0013228023890405893
step: 60, loss: 0.0005932203494012356
step: 70, loss: 0.00024350319290533662
step: 80, loss: 0.030279507860541344
step: 90, loss: 0.002330236602574587
step: 100, loss: 0.01711350493133068
step: 110, loss: 0.0002344490640098229
step: 120, loss: 0.05144970118999481
step: 130, loss: 0.0003587986866477877
step: 140, loss: 0.04588722810149193
step: 150, loss: 0.0002791484585031867
step: 160, loss: 0.0016178536461666226
step: 170, loss: 0.0002917477395385504
step: 180, loss: 0.001205048756673932
step: 190, loss: 8.599865395808592e-05
step: 200, loss: 0.00018950662342831492
step: 210, loss: 0.0001469523849664256
step: 220, loss: 0.04217581823468208
step: 230, loss: 0.012166997417807579
/usr/local/lib/python3.7/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=0.9808773903262092, f1=0.9787709497206705,
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: 0.00012110964598832652
step: 10, loss: 0.00021577607549261302
step: 20, loss: 0.00037702335976064205
step: 30, loss: 0.00017428903083782643
step: 40, loss: 8.821804658509791e-05
step: 50, loss: 0.0007950832368806005
step: 60, loss: 0.0669296383857727
step: 70, loss: 0.00019132606394123286
step: 80, loss: 0.0001385572977596894
step: 90, loss: 0.00015690881991758943
step: 100, loss: 0.0010826534125953913

```

```

step: 110, loss: 0.00021013920195400715
step: 120, loss: 0.01699107326567173
step: 130, loss: 0.0014367317780852318
step: 140, loss: 0.02023482136428356
step: 150, loss: 0.00012899124703835696
step: 160, loss: 0.000133675173856318
step: 170, loss: 0.0001744948822306469
step: 180, loss: 0.0031640841625630856
step: 190, loss: 0.0007091728621162474
step: 200, loss: 7.548864232376218e-05
step: 210, loss: 0.00016861311451066285
step: 220, loss: 0.03228650242090225
step: 230, loss: 0.0002329701092094183
/usr/local/lib/python3.7/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=0.9808342728297633, f1=0.9743016759776536,
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: 0.00010284461313858628
step: 10, loss: 0.00016025819058995694
step: 20, loss: 5.269912435323931e-05
step: 30, loss: 0.000142498713103123
step: 40, loss: 0.02514229714870453
step: 50, loss: 0.0011554218363016844
step: 60, loss: 0.0011217236751690507
step: 70, loss: 0.00013672416389454156
step: 80, loss: 0.00044053790043108165
step: 90, loss: 0.0014925834257155657
step: 100, loss: 0.0002949679037556052
step: 110, loss: 0.009272314608097076
step: 120, loss: 0.0003484167391434312
step: 130, loss: 0.00012152789713582024
step: 140, loss: 0.0001325274643022567
step: 150, loss: 9.474217222305015e-05
step: 160, loss: 0.0007109447615221143
step: 170, loss: 0.005047137849032879
step: 180, loss: 0.0002936199598480016
step: 190, loss: 0.00028903406928293407
step: 200, loss: 0.0003102862974628806
step: 210, loss: 5.869457891094498e-05
step: 220, loss: 0.0005102469585835934
step: 230, loss: 0.005029122810810804
/usr/local/lib/python3.7/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=0.9796380090497738, f1=0.9696287964004499,
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: 0.00045339815551415086
step: 10, loss: 0.0019588239956647158
step: 20, loss: 0.00032106763683259487
step: 30, loss: 0.007473325822502375
step: 40, loss: 4.664840889745392e-05
step: 50, loss: 6.629239214817062e-05
step: 60, loss: 0.006959813181310892
step: 70, loss: 7.588060543639585e-05
step: 80, loss: 4.281208020984195e-05
step: 90, loss: 8.817248453851789e-05
step: 100, loss: 0.00041461517685092986
step: 110, loss: 0.00021765779820270836
step: 120, loss: 0.0032121059484779835
step: 130, loss: 0.0001043624070007354
step: 140, loss: 0.00013922063226345927
step: 150, loss: 0.00017963594291359186
step: 160, loss: 0.019002269953489304
step: 170, loss: 6.183848017826676e-05
step: 180, loss: 5.887238512514159e-05
step: 190, loss: 5.691974001820199e-05
step: 200, loss: 0.001276654191315174
step: 210, loss: 5.042256088927388e-05
step: 220, loss: 0.018544025719165802
step: 230, loss: 5.578669879469089e-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=0.9787234042553192, f1=0.9732739420935412,
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: 0.0007075222092680633
step: 10, loss: 0.0003950854588765651
step: 20, loss: 0.00014481307880487293
step: 30, loss: 0.019424764439463615
step: 40, loss: 0.0002271544508403167
step: 50, loss: 8.607233030488715e-05
step: 60, loss: 5.451909964904189e-05
step: 70, loss: 5.959945701761171e-05
step: 80, loss: 0.0011455263011157513
step: 90, loss: 4.6137283788993955e-05
step: 100, loss: 0.00010765865590656176

```

```

step: 110, loss: 0.000122723649838008
step: 120, loss: 0.00028559198835864663
step: 130, loss: 0.007804396562278271
step: 140, loss: 0.0067654941231012344
step: 150, loss: 0.010344154201447964
step: 160, loss: 9.762221452547237e-05
step: 170, loss: 0.00016957908519543707
step: 180, loss: 0.00013157144712749869
step: 190, loss: 0.0001119705120800063
step: 200, loss: 0.00012318853987380862
step: 210, loss: 0.00013280681741889566
step: 220, loss: 5.467404116643593e-05
step: 230, loss: 5.2003746532136574e-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=0.978675645342312, f1=0.9753914988814317,
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: 8.018439257284626e-05
step: 10, loss: 3.651726001407951e-05
step: 20, loss: 0.014707008376717567
step: 30, loss: 0.00014306974480859935
step: 40, loss: 3.546261359588243e-05
step: 50, loss: 0.00038147068698890507
step: 60, loss: 3.687759453896433e-05
step: 70, loss: 0.0002735398884397
step: 80, loss: 7.870194531278685e-05
step: 90, loss: 0.00017671423847787082
step: 100, loss: 0.0065678865648806095
step: 110, loss: 0.00011385566176613793
step: 120, loss: 0.0006699724472127855
step: 130, loss: 0.0006891405209898949
step: 140, loss: 4.575429920805618e-05
step: 150, loss: 7.535119948443025e-05
step: 160, loss: 3.844175080303103e-05
step: 170, loss: 8.803373930277303e-05
step: 180, loss: 0.00016666958981659263
step: 190, loss: 0.006546509452164173
step: 200, loss: 4.54491819255054e-05
step: 210, loss: 0.015548820607364178
step: 220, loss: 0.00014145151362754405
step: 230, loss: 2.8083848519599997e-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=0.979591836734694, f1=0.9729119638826186,
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: 4.4569180317921564e-05
step: 10, loss: 0.00010530740109970793
step: 20, loss: 0.00010526335972826928
step: 30, loss: 0.00034266995498910546
step: 40, loss: 9.010918438434601e-05
step: 50, loss: 8.778279880061746e-05
step: 60, loss: 3.7477271689567715e-05
step: 70, loss: 0.0005244911881163716
step: 80, loss: 0.0010209855390712619
step: 90, loss: 3.971186015405692e-05
step: 100, loss: 8.175538823707029e-05
step: 110, loss: 4.935282777296379e-05
step: 120, loss: 6.120850594015792e-05
step: 130, loss: 6.696975469822064e-05
step: 140, loss: 0.0001502182421972975
step: 150, loss: 0.00014433058095164597
step: 160, loss: 4.3963005737168714e-05
step: 170, loss: 3.919171285815537e-05
step: 180, loss: 5.4008651204640046e-05
step: 190, loss: 0.0012746343854814768
step: 200, loss: 9.364840661874041e-05
step: 210, loss: 0.00039780032238923013
step: 220, loss: 0.0039897519163787365
step: 230, loss: 7.020151679171249e-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=0.9797752808988766, f1=0.9753914988814317,
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 [00:08, 296.69it/s]
load_f1 = 0.9808773903262092
real_f1 = 0.9797752808988766
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 [00:11, 368.95it/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/
```

```
Downloading: 100% 442/442 [00:00<00:00, 523kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 1.22MB/s]
Downloading: 100% 268M/268M [00:13<00:00, 19.2MB/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: 0.7953653335571289
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4500696659088135
step: 20, loss: 0.4936882257461548
step: 30, loss: 0.41154715418815613
step: 40, loss: 0.3031705319881439
step: 50, loss: 0.15944859385490417
step: 60, loss: 0.09556473791599274
step: 70, loss: 0.06396210938692093
step: 80, loss: 0.13818883895874023
step: 90, loss: 0.1551658809185028
step: 100, loss: 0.23812928795814514
step: 110, loss: 0.05006156861782074
step: 120, loss: 0.035687025636434555
step: 130, loss: 0.027895843610167503
step: 140, loss: 0.24168990552425385
step: 150, loss: 0.028300773352384567
step: 160, loss: 0.20872865617275238
step: 170, loss: 0.044540826231241226
step: 180, loss: 0.09638916701078415
step: 190, loss: 0.029790550470352173
step: 200, loss: 0.1278243511915207
step: 210, loss: 0.03263359144330025
step: 220, loss: 0.12038537114858627
step: 230, loss: 0.06415116041898727
step: 240, loss: 0.03198809549212456
step: 250, loss: 0.05442943051457405
step: 260, loss: 0.019289590418338776
step: 270, loss: 0.01162058673799038
step: 280, loss: 0.06042034551501274
step: 290, loss: 0.03281474485993385
step: 300, loss: 0.15935824811458588
step: 310, loss: 0.03522547706961632
step: 320, loss: 0.03357711061835289
step: 330, loss: 0.09950681030750275
step: 340, loss: 0.19026988744735718
step: 350, loss: 0.05632374808192253
step: 360, loss: 0.04438957944512367
step: 370, loss: 0.1590936928987503
step: 380, loss: 0.18390366435050964
step: 390, loss: 0.018181271851062775
step: 400, loss: 0.010359102860093117
step: 410, loss: 0.051026660948991776
step: 420, loss: 0.009272273629903793
step: 430, loss: 0.06553685665130615

```

```

step: 440, loss: 0.08112279325723648
step: 450, loss: 0.03864305093884468
step: 460, loss: 0.06791822612285614
step: 470, loss: 0.13004711270332336
step: 480, loss: 0.301601380109787
step: 490, loss: 0.01660769432783127
step: 500, loss: 0.00533848162740469
step: 510, loss: 0.01950569823384285
step: 520, loss: 0.08137564361095428
step: 530, loss: 0.15158838033676147
/usr/local/lib/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=0.9301675977653632, f1=0.9302973977695168,
best_f1=0.9302973977695168
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0685667023062706
step: 10, loss: 0.15483225882053375
step: 20, loss: 0.09386665374040604
step: 30, loss: 0.008341091684997082
step: 40, loss: 0.004398851655423641
step: 50, loss: 0.05788644030690193
step: 60, loss: 0.0877198576927185
step: 70, loss: 0.1173001378774643
step: 80, loss: 0.01415393128991127
step: 90, loss: 0.0009927062783390284
step: 100, loss: 0.21879985928535461
step: 110, loss: 0.011693285778164864
step: 120, loss: 0.11809252947568893
step: 130, loss: 0.025480056181550026
step: 140, loss: 0.05167488008737564
step: 150, loss: 0.0881248414516449
step: 160, loss: 0.08929122984409332
step: 170, loss: 0.19711174070835114
step: 180, loss: 0.014661027118563652
step: 190, loss: 0.016791442409157753
step: 200, loss: 0.009118271991610527
step: 210, loss: 0.005877605173736811
step: 220, loss: 0.14620091021060944
step: 230, loss: 0.03431867063045502
step: 240, loss: 0.1103280782699585
step: 250, loss: 0.019817350432276726

```

```

step: 260, loss: 0.045520614832639694
step: 270, loss: 0.03603430092334747
step: 280, loss: 0.09175832569599152
step: 290, loss: 0.06292008608579636
step: 300, loss: 0.009059385396540165
step: 310, loss: 0.04274482652544975
step: 320, loss: 0.13240021467208862
step: 330, loss: 0.006780573166906834
step: 340, loss: 0.09290270507335663
step: 350, loss: 0.04912286996841431
step: 360, loss: 0.02364184334874153
step: 370, loss: 0.018877403810620308
step: 380, loss: 0.04591265320777893
step: 390, loss: 0.0329422764480114
step: 400, loss: 0.024941934272646904
step: 410, loss: 0.0007065749377943575
step: 420, loss: 0.06236502528190613
step: 430, loss: 0.02666236087679863
step: 440, loss: 0.013704788871109486
step: 450, loss: 0.008035758510231972
step: 460, loss: 0.22491800785064697
step: 470, loss: 0.025399718433618546
step: 480, loss: 0.17960260808467865
step: 490, loss: 0.014517104253172874
step: 500, loss: 0.00907948799431324
step: 510, loss: 0.02769327536225319
step: 520, loss: 0.1571832001209259
step: 530, loss: 0.09063874185085297
/usr/local/lib/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=0.945736434108527, f1=0.9378684807256237,
best_f1=0.9378684807256237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014326183125376701
step: 10, loss: 0.09480945765972137
step: 20, loss: 0.09502018988132477
step: 30, loss: 0.2210616022348404
step: 40, loss: 0.004591385833919048
step: 50, loss: 0.019018717110157013
step: 60, loss: 0.017875371500849724
step: 70, loss: 0.05174875631928444

```

step: 80, loss: 0.0033583606127649546
step: 90, loss: 0.03658628091216087
step: 100, loss: 0.04455700144171715
step: 110, loss: 0.006143040023744106
step: 120, loss: 0.007644266355782747
step: 130, loss: 0.024317951872944832
step: 140, loss: 0.024704642593860626
step: 150, loss: 0.0026705858763307333
step: 160, loss: 0.01346883736550808
step: 170, loss: 0.014670299366116524
step: 180, loss: 0.020194094628095627
step: 190, loss: 0.005817107390612364
step: 200, loss: 0.007890848442912102
step: 210, loss: 0.0751783549785614
step: 220, loss: 0.009631996043026447
step: 230, loss: 0.059109464287757874
step: 240, loss: 0.021128984168171883
step: 250, loss: 0.0214361771941185
step: 260, loss: 0.0032049138098955154
step: 270, loss: 0.0017778389155864716
step: 280, loss: 0.0046150824055075645
step: 290, loss: 0.021635035052895546
step: 300, loss: 0.0864589586853981
step: 310, loss: 0.09814272820949554
step: 320, loss: 0.15457157790660858
step: 330, loss: 0.02036173641681671
step: 340, loss: 0.0185188390314579
step: 350, loss: 0.11063414067029953
step: 360, loss: 0.016990168020129204
step: 370, loss: 0.0013367378851398826
step: 380, loss: 0.007819361053407192
step: 390, loss: 0.08536789566278458
step: 400, loss: 0.0513351634144783
step: 410, loss: 0.015364578925073147
step: 420, loss: 0.012428433634340763
step: 430, loss: 0.016551081091165543
step: 440, loss: 0.02422342076897621
step: 450, loss: 0.16666993498802185
step: 460, loss: 0.07571637630462646
step: 470, loss: 0.004340847488492727
step: 480, loss: 0.006091400980949402
step: 490, loss: 0.015403770841658115
step: 500, loss: 0.05429453030228615
step: 510, loss: 0.001542729209177196
step: 520, loss: 0.011841259896755219
step: 530, loss: 0.031202683225274086
/usr/local/lib/python3.7/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=0.9412861136999067, f1=0.942271880819367,
best_f1=0.9378684807256237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013996890746057034
step: 10, loss: 0.01401604525744915
step: 20, loss: 0.03600824996829033
step: 30, loss: 0.0030556584242731333
step: 40, loss: 0.007689951919019222
step: 50, loss: 0.015656426548957825
step: 60, loss: 0.002097448566928506
step: 70, loss: 0.004070964176207781
step: 80, loss: 0.06300836056470871
step: 90, loss: 0.10580886900424957
step: 100, loss: 0.025967329740524292
step: 110, loss: 0.00250052148476243
step: 120, loss: 0.0015822014538571239
step: 130, loss: 0.13592790067195892
step: 140, loss: 0.01607789285480976
step: 150, loss: 0.008718865923583508
step: 160, loss: 0.016040699556469917
step: 170, loss: 0.023098278790712357
step: 180, loss: 0.03861559182405472
step: 190, loss: 0.004024439956992865
step: 200, loss: 0.0039020669646561146
step: 210, loss: 0.07487594336271286
step: 220, loss: 0.009487343952059746
step: 230, loss: 0.2648794949054718
step: 240, loss: 0.012691725976765156
step: 250, loss: 0.030424779281020164
step: 260, loss: 0.12103935331106186
step: 270, loss: 0.13194730877876282
step: 280, loss: 0.004407917149364948
step: 290, loss: 0.0022332954686135054
step: 300, loss: 0.008508739061653614
step: 310, loss: 0.0015147635713219643
step: 320, loss: 0.0023856505285948515
step: 330, loss: 0.0491085909307003
step: 340, loss: 0.02579895220696926
step: 350, loss: 0.004687601700425148
step: 360, loss: 0.00472386134788394
step: 370, loss: 0.11519548296928406
step: 380, loss: 0.006189856678247452
step: 390, loss: 0.00697902450338006
step: 400, loss: 0.006890888325870037
step: 410, loss: 0.05153396725654602
```

```

step: 420, loss: 0.0006651650764979422
step: 430, loss: 0.02813383750617504
step: 440, loss: 0.07495611160993576
step: 450, loss: 0.025164835155010223
step: 460, loss: 0.044925056397914886
step: 470, loss: 0.012177983298897743
step: 480, loss: 0.0020445699337869883
step: 490, loss: 0.03676271438598633
step: 500, loss: 0.009924921207129955
step: 510, loss: 0.02712324820458889
step: 520, loss: 0.033484991639852524
step: 530, loss: 0.0015096693532541394
/usr/local/lib/python3.7/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=0.9444184960298926, f1=0.9430438842203548,
best_f1=0.9378684807256237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00228859088383615
step: 10, loss: 0.0031523110810667276
step: 20, loss: 0.000887452915776521
step: 30, loss: 0.0015560828614979982
step: 40, loss: 0.03976615518331528
step: 50, loss: 0.005573122296482325
step: 60, loss: 0.0033753139432519674
step: 70, loss: 0.003049329621717334
step: 80, loss: 0.00455815251916647
step: 90, loss: 0.10247120261192322
step: 100, loss: 0.0010308941127732396
step: 110, loss: 0.050861652940511703
step: 120, loss: 0.009604752995073795
step: 130, loss: 0.000726801052223891
step: 140, loss: 0.00023124607105273753
step: 150, loss: 0.001969381934031844
step: 160, loss: 0.057410724461078644
step: 170, loss: 0.061956536024808884
step: 180, loss: 0.0005192364333197474
step: 190, loss: 0.012652460485696793
step: 200, loss: 0.0027874575462192297
step: 210, loss: 0.00399682717397809
step: 220, loss: 0.0013181426329538226
step: 230, loss: 0.0014430010924115777
step: 240, loss: 0.01435309648513794
step: 250, loss: 0.0006688625435344875
step: 260, loss: 0.0014321248745545745
step: 270, loss: 0.13554716110229492

```

```

step: 280, loss: 0.0469331294298172
step: 290, loss: 0.004920742474496365
step: 300, loss: 0.038466647267341614
step: 310, loss: 0.0006212841835804284
step: 320, loss: 0.022090669721364975
step: 330, loss: 0.0005986507749184966
step: 340, loss: 0.0009193555451929569
step: 350, loss: 0.0004964646650478244
step: 360, loss: 0.05712650343775749
step: 370, loss: 0.0013021520571783185
step: 380, loss: 0.025684446096420288
step: 390, loss: 0.03953971713781357
step: 400, loss: 0.07243695110082626
step: 410, loss: 0.006704403553158045
step: 420, loss: 0.0009270681766793132
step: 430, loss: 0.060512833297252655
step: 440, loss: 0.002693016780540347
step: 450, loss: 0.009776820428669453
step: 460, loss: 0.032930463552474976
step: 470, loss: 0.0020106404554098845
step: 480, loss: 0.001523106824606657
step: 490, loss: 0.05271776020526886
step: 500, loss: 0.040931686758995056
step: 510, loss: 0.014809731394052505
step: 520, loss: 0.00413523381575942
step: 530, loss: 0.0024559968151152134
/usr/local/lib/python3.7/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=0.943219145940873, f1=0.9395218002812938,
best_f1=0.9378684807256237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008730254136025906
step: 10, loss: 0.00247387308627367
step: 20, loss: 0.0009476139093749225
step: 30, loss: 0.0030805973801761866
step: 40, loss: 0.005376526620239019
step: 50, loss: 0.0026975288055837154
step: 60, loss: 0.02209807187318802
step: 70, loss: 0.005786826368421316
step: 80, loss: 0.02464420162141323
step: 90, loss: 0.0016134128673002124
step: 100, loss: 0.016295399516820908
step: 110, loss: 0.01136729959398508
step: 120, loss: 0.02943844348192215
step: 130, loss: 0.000652445771265775

```



```

step: 140, loss: 0.0005316011956892908
step: 150, loss: 0.0008127741166390479
step: 160, loss: 0.0014072187477722764
step: 170, loss: 0.0029999203979969025
step: 180, loss: 0.002868352923542261
step: 190, loss: 0.025253374129533768
step: 200, loss: 0.0005569085478782654
step: 210, loss: 0.0017186475452035666
step: 220, loss: 0.0017322067869827151
step: 230, loss: 0.0017752008279785514
step: 240, loss: 0.0006248687859624624
step: 250, loss: 0.003542541293427348
step: 260, loss: 0.0006969473324716091
step: 270, loss: 0.002402088139206171
step: 280, loss: 0.0024060753639787436
step: 290, loss: 0.0010575932683423162
step: 300, loss: 0.007718429435044527
step: 310, loss: 0.0002747366379480809
step: 320, loss: 0.041337013244628906
step: 330, loss: 0.0042608678340911865
step: 340, loss: 0.0022435584105551243
step: 350, loss: 0.11942388117313385
step: 360, loss: 0.014173480682075024
step: 370, loss: 0.003424927359446883
step: 380, loss: 0.0033280649222433567
step: 390, loss: 0.0018082137685269117
step: 400, loss: 0.002051545772701502
step: 410, loss: 0.00030674785375595093
step: 420, loss: 0.05508805811405182
step: 430, loss: 0.0002450921747367829
step: 440, loss: 0.02585931122303009
step: 450, loss: 0.001311193685978651
step: 460, loss: 0.003990212921053171
step: 470, loss: 0.013271176256239414
step: 480, loss: 0.004828604403883219
step: 490, loss: 0.01884416490793228
step: 500, loss: 0.007587885949760675
step: 510, loss: 0.0016233851201832294
step: 520, loss: 0.009952602908015251
step: 530, loss: 0.0018780251266434789
/usr/local/lib/python3.7/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=0.9431500465983226, f1=0.9386617100371748,
best_f1=0.9378684807256237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.023095883429050446
step: 10, loss: 0.0006679670186713338
step: 20, loss: 0.0071318442933261395
step: 30, loss: 0.009786759503185749
step: 40, loss: 0.0014602919109165668
step: 50, loss: 0.0014342068461701274
step: 60, loss: 0.04941144213080406
step: 70, loss: 0.027081001549959183
step: 80, loss: 0.0007978870999068022
step: 90, loss: 0.00036743978853337467
step: 100, loss: 0.0016512228175997734
step: 110, loss: 0.0020254801493138075
step: 120, loss: 0.005150475073605776
step: 130, loss: 0.002480152528733015
step: 140, loss: 0.013028002344071865
step: 150, loss: 0.0002374895993852988
step: 160, loss: 0.014644436538219452
step: 170, loss: 0.002238561399281025
step: 180, loss: 0.0018638904439285398
step: 190, loss: 0.00016615305503364652
step: 200, loss: 0.002180578652769327
step: 210, loss: 0.004359431564807892
step: 220, loss: 0.00013970279542263597
step: 230, loss: 0.0027417833916842937
step: 240, loss: 0.0030226048547774553
step: 250, loss: 0.0012099901214241982
step: 260, loss: 0.0041632093489170074
step: 270, loss: 0.006756218150258064
step: 280, loss: 0.006650789640843868
step: 290, loss: 0.003563017351552844
step: 300, loss: 0.0025711683556437492
step: 310, loss: 0.0001254638482350856
step: 320, loss: 0.06426391005516052
step: 330, loss: 0.003966838587075472
step: 340, loss: 0.0034386429470032454
step: 350, loss: 0.11376215517520905
step: 360, loss: 0.0023350766859948635
step: 370, loss: 0.0024017926771193743
step: 380, loss: 0.003402580274268985
step: 390, loss: 0.00012764126586262137
step: 400, loss: 0.0007029462140053511
step: 410, loss: 0.0024457229301333427
step: 420, loss: 0.005191086325794458
step: 430, loss: 0.00029161712154746056
step: 440, loss: 0.002582875546067953
step: 450, loss: 0.006178764626383781
step: 460, loss: 0.009590757079422474
step: 470, loss: 0.06343907117843628

```

step: 480, loss: 0.012779065407812595
step: 490, loss: 0.0016870342660695314
step: 500, loss: 0.0023368322290480137
step: 510, loss: 0.011717363260686398
step: 520, loss: 0.0011153389932587743
step: 530, loss: 9.370715997647494e-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 7: dev_f1=0.9483480688692415, f1=0.9397031539888682,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0033516765106469393
step: 10, loss: 0.005705049727112055
step: 20, loss: 0.038658346980810165
step: 30, loss: 0.006605965085327625
step: 40, loss: 0.0036502659786492586
step: 50, loss: 0.00047219969565048814
step: 60, loss: 0.0059591056779026985
step: 70, loss: 0.0005151384393684566
step: 80, loss: 0.00014845251280348748
step: 90, loss: 0.0003438152198214084
step: 100, loss: 0.05054723843932152
step: 110, loss: 0.0004021405184175819
step: 120, loss: 0.002291366457939148
step: 130, loss: 0.010560993105173111
step: 140, loss: 0.00018886731413658708
step: 150, loss: 0.001429478288628161
step: 160, loss: 0.005780717357993126
step: 170, loss: 0.04838278144598007
step: 180, loss: 6.513184780487791e-05
step: 190, loss: 0.003182763233780861
step: 200, loss: 0.00024308347201440483
step: 210, loss: 0.0014004816766828299
step: 220, loss: 0.0009195465827360749
step: 230, loss: 0.00012628089461941272
step: 240, loss: 0.0009010043577291071
step: 250, loss: 0.00026392636937089264
step: 260, loss: 0.028154799714684486
step: 270, loss: 8.237740985350683e-05
step: 280, loss: 0.007626195903867483
step: 290, loss: 0.004830304533243179

```

```

step: 300, loss: 0.0001783974439604208
step: 310, loss: 0.06478332728147507
step: 320, loss: 0.0003936404245905578
step: 330, loss: 0.01133892871439457
step: 340, loss: 0.0015420997515320778
step: 350, loss: 0.02039257064461708
step: 360, loss: 0.0003427263582125306
step: 370, loss: 0.0009467334020882845
step: 380, loss: 0.010797268711030483
step: 390, loss: 0.010814777575433254
step: 400, loss: 0.2518830895423889
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 410, loss: 0.00013261022104416043
step: 420, loss: 0.0006873221136629581
step: 430, loss: 0.0026551426853984594
step: 440, loss: 0.0022839016746729612
step: 450, loss: 0.00036129806539975107
step: 460, loss: 0.0002922996645793319
step: 470, loss: 0.006655754987150431
step: 480, loss: 0.00027854001382365823
step: 490, loss: 9.947154467226937e-05
step: 500, loss: 0.00031422567553818226
step: 510, loss: 0.009061699733138084
step: 520, loss: 0.0003335383371450007
step: 530, loss: 0.014859460294246674
/usr/local/lib/python3.7/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=0.947906976744186, f1=0.937471051412691,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028422035393305123
step: 10, loss: 0.00020604029123205692
step: 20, loss: 0.00020360264170449227
step: 30, loss: 0.0002492183120921254
step: 40, loss: 0.018270395696163177
step: 50, loss: 0.004803772084414959
step: 60, loss: 0.000499464338645339
step: 70, loss: 0.12788377702236176
step: 80, loss: 0.0022317247930914164
step: 90, loss: 0.022277778014540672
step: 100, loss: 0.001927760778926313
step: 110, loss: 0.0019791938830167055
step: 120, loss: 0.0002387093409197405
step: 130, loss: 0.008506344631314278
step: 140, loss: 0.0004040436469949782

```

```

step: 150, loss: 3.1291529012378305e-05
step: 160, loss: 2.388986285950523e-05
step: 170, loss: 0.0004909925046376884
step: 180, loss: 3.011793342011515e-05
step: 190, loss: 0.0006988546811044216
step: 200, loss: 0.00018614722648635507
step: 210, loss: 0.00016024455544538796
step: 220, loss: 0.0005445131682790816
step: 230, loss: 8.984583837445825e-05
step: 240, loss: 5.9025653172284365e-05
step: 250, loss: 0.00018693789024837315
step: 260, loss: 7.79778347350657e-05
step: 270, loss: 0.0006013885140419006
step: 280, loss: 2.9730828828178346e-05
step: 290, loss: 3.069535523536615e-05
step: 300, loss: 0.0001965608389582485
step: 310, loss: 2.0727271476062015e-05
step: 320, loss: 0.00041706347838044167
step: 330, loss: 0.0011118942638859153
step: 340, loss: 5.459912063088268e-05
step: 350, loss: 3.5407269024290144e-05
step: 360, loss: 0.018699385225772858
step: 370, loss: 0.00039897506940178573
step: 380, loss: 2.7127083740197122e-05
step: 390, loss: 4.8941954446490854e-05
step: 400, loss: 0.09275974333286285
step: 410, loss: 0.0001288453204324469
step: 420, loss: 0.0005439411615952849
step: 430, loss: 5.0853941502282396e-05
step: 440, loss: 0.019465912133455276
step: 450, loss: 0.014802700839936733
step: 460, loss: 0.0006801748531870544
step: 470, loss: 2.2876609364175238e-05
step: 480, loss: 0.00010180650133406743
step: 490, loss: 9.29918242036365e-05
step: 500, loss: 0.06359483301639557
step: 510, loss: 0.005117433145642281
step: 520, loss: 0.004891528282314539
step: 530, loss: 0.00012122482439735904
/usr/local/lib/python3.7/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=0.9461573861021628, f1=0.9381584974805315,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024233919102698565

```

step: 10, loss: 3.4751570638036355e-05
step: 20, loss: 0.0016541184158995748
step: 30, loss: 8.013626938918605e-05
step: 40, loss: 0.005162738263607025
step: 50, loss: 0.0004286030598450452
step: 60, loss: 0.00021406926680356264
step: 70, loss: 0.059440728276968
step: 80, loss: 0.0007020154152996838
step: 90, loss: 0.0002592981909401715
step: 100, loss: 0.00022877665469422936
step: 110, loss: 0.0004163295670878142
step: 120, loss: 0.0004994868068024516
step: 130, loss: 0.02013939991593361
step: 140, loss: 0.0008716183947399259
step: 150, loss: 9.087654325412586e-05
step: 160, loss: 0.0005016414797864854
step: 170, loss: 0.03003089502453804
step: 180, loss: 0.0010858624009415507
step: 190, loss: 0.0022567706182599068
step: 200, loss: 0.0010148199507966638
step: 210, loss: 0.0022178133949637413
step: 220, loss: 0.008061308413743973
step: 230, loss: 0.00011085345613537356
step: 240, loss: 1.729257928673178e-05
step: 250, loss: 0.0008097745012491941
step: 260, loss: 0.0015484769828617573
step: 270, loss: 0.011695810593664646
step: 280, loss: 0.0023989868350327015
step: 290, loss: 0.00035216580727137625
step: 300, loss: 0.007853846065700054
step: 310, loss: 0.0005684021161869168
step: 320, loss: 0.0006076885620132089
step: 330, loss: 0.004685006104409695
step: 340, loss: 0.0006523655029013753
step: 350, loss: 0.0003244881227146834
step: 360, loss: 0.00021105416817590594
step: 370, loss: 0.006035737693309784
step: 380, loss: 0.007383933290839195
step: 390, loss: 7.176835060818121e-05
step: 400, loss: 0.008190386928617954
step: 410, loss: 0.00031296355882659554
step: 420, loss: 0.006131108850240707
step: 430, loss: 0.00047140903188847005
step: 440, loss: 2.4563785700593144e-05
step: 450, loss: 0.014519303105771542
step: 460, loss: 0.016904456540942192
step: 470, loss: 0.0002545367751736194
step: 480, loss: 0.00042659323662519455

```
step: 490, loss: 0.05864899978041649
step: 500, loss: 0.01086751464754343
step: 510, loss: 0.02547876536846161
step: 520, loss: 0.0038620620034635067
step: 530, loss: 0.00021422393911052495
/usr/local/lib/python3.7/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=0.9423697556477639, f1=0.9412304866850323,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006619682069867849
step: 10, loss: 0.0641731396317482
step: 20, loss: 0.002519250148907304
step: 30, loss: 0.006021346431225538
step: 40, loss: 0.0033740480430424213
step: 50, loss: 0.004257886670529842
step: 60, loss: 0.00021245115203782916
step: 70, loss: 9.48402812355198e-05
step: 80, loss: 0.00788182020187378
step: 90, loss: 0.0008172263042069972
step: 100, loss: 0.0012497749412432313
step: 110, loss: 0.00014775320596527308
step: 120, loss: 0.0010667848400771618
step: 130, loss: 0.0014031059108674526
step: 140, loss: 0.0011504964204505086
step: 150, loss: 0.000833365018479526
step: 160, loss: 0.005099070258438587
step: 170, loss: 0.0023218300193548203
step: 180, loss: 0.00031695447978563607
step: 190, loss: 0.0003672951424960047
step: 200, loss: 0.13506294786930084
step: 210, loss: 9.515426063444465e-05
step: 220, loss: 0.0015029993373900652
step: 230, loss: 0.001081557129509747
step: 240, loss: 0.0009801493724808097
step: 250, loss: 0.001700564636848867
step: 260, loss: 0.0016207625158131123
step: 270, loss: 0.0014177619013935328
step: 280, loss: 0.004322819411754608
step: 290, loss: 0.0015001524006947875
step: 300, loss: 0.005980649497359991
step: 310, loss: 0.00046402931911870837
step: 320, loss: 0.03285350650548935
step: 330, loss: 0.0029894718900322914
step: 340, loss: 0.001390055171214044
```

```

step: 350, loss: 0.0018613676074892282
step: 360, loss: 0.0013232608325779438
step: 370, loss: 0.0006325674476101995
step: 380, loss: 0.0002514233347028494
step: 390, loss: 0.19897207617759705
step: 400, loss: 0.00014485139399766922
step: 410, loss: 0.001813627197407186
step: 420, loss: 0.0019244300201535225
step: 430, loss: 0.0729391947388649
step: 440, loss: 0.0017374262679368258
step: 450, loss: 0.00040174368768930435
step: 460, loss: 0.0016053543658927083
step: 470, loss: 0.0005160779110156
step: 480, loss: 0.0013480732450261712
step: 490, loss: 0.012724745087325573
step: 500, loss: 0.00036912807263433933
step: 510, loss: 0.0004019223270006478
step: 520, loss: 3.6634490243159235e-05
step: 530, loss: 4.2621988541213796e-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=0.947906976744186, f1=0.9415041782729804,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008588763885200024
step: 10, loss: 0.005008217878639698
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 7.323649333557114e-05
step: 30, loss: 0.0001947414129972458
step: 40, loss: 0.0003496516728773713
step: 50, loss: 0.00035662815207615495
step: 60, loss: 8.585195610066876e-05
step: 70, loss: 0.0011648706858977675
step: 80, loss: 0.0007256101816892624
step: 90, loss: 0.032792989164590836
step: 100, loss: 0.0035347389057278633
step: 110, loss: 0.0008678134763613343
step: 120, loss: 0.0006839020643383265
step: 130, loss: 0.003645116463303566
step: 140, loss: 0.00012856200919486582
step: 150, loss: 0.0009275984484702349
step: 160, loss: 0.008130036294460297
step: 170, loss: 0.0008887371513992548
step: 180, loss: 0.0026084345299750566
step: 190, loss: 5.944617805653252e-05

```



```
step: 200, loss: 2.8914702852489427e-05
step: 210, loss: 0.00014934710634406656
step: 220, loss: 2.4086295525194146e-05
step: 230, loss: 0.0038458590861409903
step: 240, loss: 5.358047201298177e-05
step: 250, loss: 0.0016608776059001684
step: 260, loss: 8.927703311201185e-05
step: 270, loss: 4.047858965350315e-05
step: 280, loss: 0.00083530368283391
step: 290, loss: 0.008596368134021759
step: 300, loss: 0.0009884900646284223
step: 310, loss: 1.7661179299466312e-05
step: 320, loss: 2.934571966761723e-05
step: 330, loss: 4.9555055738892406e-05
step: 340, loss: 0.0027075367979705334
step: 350, loss: 0.015484346076846123
step: 360, loss: 0.0016423023771494627
step: 370, loss: 5.89176343055442e-05
step: 380, loss: 0.0011984746670350432
step: 390, loss: 0.000979039235971868
step: 400, loss: 0.0010341275483369827
step: 410, loss: 0.004064745269715786
step: 420, loss: 0.0013829405652359128
step: 430, loss: 0.004610739182680845
step: 440, loss: 0.0003386458847671747
step: 450, loss: 0.003436560742557049
step: 460, loss: 0.016194632276892662
step: 470, loss: 0.00010273468069499359
step: 480, loss: 0.0005381755763664842
step: 490, loss: 0.0024858927354216576
step: 500, loss: 0.004984931088984013
step: 510, loss: 0.0002702563360799104
step: 520, loss: 0.023002756759524345
step: 530, loss: 0.00014893860497977585
/usr/local/lib/python3.7/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=0.9477093937991671, f1=0.9428044280442804,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001037150272168219
step: 10, loss: 0.0002567195915617049
step: 20, loss: 0.02442806400358677
step: 30, loss: 0.007298167794942856
step: 40, loss: 0.00013860156468581408
step: 50, loss: 0.00038285390473902225
```

step: 60, loss: 6.792198109906167e-05
step: 70, loss: 0.00017925792781170458
step: 80, loss: 0.000150352789205499
step: 90, loss: 0.005653129890561104
step: 100, loss: 1.8789549358189106e-05
step: 110, loss: 0.20296235382556915
step: 120, loss: 0.00012133528070989996
step: 130, loss: 0.0007440508343279362
step: 140, loss: 0.0001321111194556579
step: 150, loss: 0.0009062154567800462
step: 160, loss: 1.6577254427829757e-05
step: 170, loss: 0.0038070257287472486
step: 180, loss: 5.905690341023728e-05
step: 190, loss: 0.0003928758669644594
step: 200, loss: 0.04832311347126961
step: 210, loss: 0.00034154998138546944
step: 220, loss: 0.00017294623830821365
step: 230, loss: 0.00046756461961194873
step: 240, loss: 2.6196941689704545e-05
step: 250, loss: 0.03183048218488693
step: 260, loss: 0.00020497145305853337
step: 270, loss: 0.022404013201594353
step: 280, loss: 0.0006228170823305845
step: 290, loss: 0.00017095702060032636
step: 300, loss: 0.0003717601066455245
step: 310, loss: 0.00011277430894551799
step: 320, loss: 0.00025500074843876064
step: 330, loss: 0.00028600520454347134
step: 340, loss: 0.0003074627311434597
step: 350, loss: 0.14909209311008453
step: 360, loss: 2.6772839191835374e-05
step: 370, loss: 0.0003156384627800435
step: 380, loss: 0.0011582787847146392
step: 390, loss: 0.00016808006330393255
step: 400, loss: 0.00035322600160725415
step: 410, loss: 3.202362859155983e-05
step: 420, loss: 3.8180551200639457e-05
step: 430, loss: 0.0037953886203467846
step: 440, loss: 6.621183274546638e-05
step: 450, loss: 5.534271986107342e-05
step: 460, loss: 9.621152275940403e-05
step: 470, loss: 0.021770551800727844
step: 480, loss: 0.000361905200406909
step: 490, loss: 6.0765429225284606e-05
step: 500, loss: 4.423553036758676e-05
step: 510, loss: 0.000494934618473053
step: 520, loss: 3.23141721310094e-05
step: 530, loss: 0.00015596815501339734

```

/usr/local/lib/python3.7/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=0.947906976744186, f1=0.9431870669745959,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.047494389349595e-05
step: 10, loss: 5.10348254465498e-05
step: 20, loss: 0.0010853152489289641
step: 30, loss: 0.00011159990390297025
step: 40, loss: 0.012446463108062744
step: 50, loss: 0.00012639039778150618
step: 60, loss: 1.222258015332045e-05
step: 70, loss: 2.191503699577879e-05
step: 80, loss: 0.0002627242647577077
step: 90, loss: 0.0006797177484259009
step: 100, loss: 0.0003981876070611179
step: 110, loss: 3.343668140587397e-05
step: 120, loss: 0.0003255662741139531
step: 130, loss: 0.0002689354005269706
step: 140, loss: 0.015383853577077389
step: 150, loss: 0.0006147878593765199
step: 160, loss: 0.0038832570426166058
step: 170, loss: 7.04293925082311e-05
step: 180, loss: 0.0018595708534121513
step: 190, loss: 0.001914853579364717
step: 200, loss: 0.0003361298586241901
step: 210, loss: 0.0005881534889340401
step: 220, loss: 0.000774542277213186
step: 230, loss: 0.0006306060240603983
step: 240, loss: 6.10674251220189e-05
step: 250, loss: 0.0013567913556471467
step: 260, loss: 2.150550062651746e-05
step: 270, loss: 0.0029997825622558594
step: 280, loss: 2.7286461772746406e-05
step: 290, loss: 0.000392519315937534
step: 300, loss: 7.417573942802846e-05
step: 310, loss: 0.00018269197607878596
step: 320, loss: 0.00035410505370236933
step: 330, loss: 0.0005851282039657235
step: 340, loss: 0.00010099515930050984
step: 350, loss: 0.0030385900754481554
step: 360, loss: 0.009434395469725132
step: 370, loss: 0.008917526341974735
step: 380, loss: 1.9210827304050326e-05
step: 390, loss: 0.00041357576264999807

```

```

step: 400, loss: 4.625558358384296e-05
step: 410, loss: 0.0006791821215301752
step: 420, loss: 0.000825928058475256
step: 430, loss: 2.0726849470520392e-05
step: 440, loss: 2.0451225282158703e-05
step: 450, loss: 8.576864638598636e-05
step: 460, loss: 0.0017665015766397119
step: 470, loss: 0.0002774186432361603
step: 480, loss: 3.8600119296461344e-05
step: 490, loss: 0.000598300714045763
step: 500, loss: 0.013101860880851746
step: 510, loss: 2.921481427620165e-05
step: 520, loss: 2.5137069314951077e-05
step: 530, loss: 0.0013890546979382634
/usr/local/lib/python3.7/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=0.9451162790697675, f1=0.9431345353675451,
best_f1=0.9397031539888682
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021218416804913431
step: 10, loss: 9.696622146293521e-05
step: 20, loss: 5.3493768064072356e-05
step: 30, loss: 9.299404337070882e-05
step: 40, loss: 3.8202731957426295e-05
step: 50, loss: 0.0005469325114972889
step: 60, loss: 3.678963548736647e-05
step: 70, loss: 0.00015955579874571413
step: 80, loss: 0.0002200917951995507
step: 90, loss: 1.3776016203337349e-05
step: 100, loss: 0.00033616489963606
step: 110, loss: 0.018591927364468575
step: 120, loss: 0.0011909774038940668
step: 130, loss: 1.855892151070293e-05
step: 140, loss: 2.457819937262684e-05
step: 150, loss: 0.0011045272694900632
step: 160, loss: 1.8127007933799177e-05
step: 170, loss: 2.0484869310166687e-05
step: 180, loss: 0.00040216633351519704
step: 190, loss: 7.348405051743612e-05
step: 200, loss: 4.337029531598091e-05
step: 210, loss: 0.00016158810467459261
step: 220, loss: 2.4079145077848807e-05
step: 230, loss: 0.007784687913954258
step: 240, loss: 4.7921654186211526e-05
step: 250, loss: 3.526790897012688e-05

```

```

step: 260, loss: 1.814559436752461e-05
step: 270, loss: 0.0004030184354633093
step: 280, loss: 0.0005457974621094763
step: 290, loss: 0.002075319644063711
step: 300, loss: 0.0010481153149157763
step: 310, loss: 0.00018249917775392532
step: 320, loss: 0.0026147121097892523
step: 330, loss: 5.830506779602729e-05
step: 340, loss: 0.0021912052761763334
step: 350, loss: 0.0014236271381378174
step: 360, loss: 0.0018825476290658116
step: 370, loss: 9.145576768787578e-05
step: 380, loss: 8.530961349606514e-05
step: 390, loss: 6.816787936259061e-05
step: 400, loss: 4.266677933628671e-05
step: 410, loss: 0.00019005087960977107
step: 420, loss: 2.5900788386934437e-05
step: 430, loss: 3.21971601806581e-05
step: 440, loss: 0.000853242352604866
step: 450, loss: 0.007193997036665678
step: 460, loss: 0.00024308846332132816
step: 470, loss: 8.406257256865501e-05
step: 480, loss: 5.859036173205823e-05
step: 490, loss: 6.418852717615664e-05
step: 500, loss: 0.007743509486317635
step: 510, loss: 0.000637321500107646
step: 520, loss: 3.8080368540249765e-05
step: 530, loss: 1.5824845831957646e-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=0.9463369108726085, f1=0.9442896935933148,
best_f1=0.9397031539888682
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 [00:15, 369.50it/s]
load_f1 = 0.9479502533394749
real_f1 = 0.9457221711131555
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 [00:12, 339.87it/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: 0.8615951538085938
/usr/local/lib/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=0.2828282828282828, f1=0.2772272272272272,
best_f1=0.2772272272272272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3568151295185089
/usr/local/lib/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=0.2857142857142857, f1=0.5, best_f1=0.5
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32503965497016907
/usr/local/lib/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=0.39215686274509803, f1=0.32786885245901637,
best_f1=0.32786885245901637
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3376908600330353
/usr/local/lib/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=0.4545454545454545, f1=0.36363636363636365,

```



```

best_f1=0.36363636363636365
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23567664623260498
/usr/local/lib/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=0.46511627906976755, f1=0.39215686274509803,
best_f1=0.39215686274509803
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21330630779266357
/usr/local/lib/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=0.5128205128205129, f1=0.36734693877551017,
best_f1=0.36734693877551017
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21571123600006104
/usr/local/lib/python3.7/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=0.423076923076923, best_f1=0.36734693877551017
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29708173871040344
/usr/local/lib/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=0.5806451612903226, f1=0.4864864864864865,
best_f1=0.4864864864864865

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14295820891857147
/usr/local/lib/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=0.6000000000000001, f1=0.4864864864864865,
best_f1=0.4864864864864865
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12660221755504608
/usr/local/lib/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=0.6666666666666666, f1=0.6206896551724138,
best_f1=0.6206896551724138
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11276543140411377
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.5625000000000001,
best_f1=0.6206896551724138
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1135575994849205
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.6206896551724138,
best_f1=0.6206896551724138
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.041730839759111404

```

```

/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.5625000000000001,
best_f1=0.6206896551724138
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07593964785337448
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.5625000000000001,
best_f1=0.6206896551724138
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1110735535621643
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.5625000000000001,
best_f1=0.6206896551724138
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, 109239.17it/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 = 0.6666666666666666
real_f1 = 0.6206896551724138
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, 270.42it/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        : 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: 0.8116968870162964

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

step: 10, loss: 0.4733388423919678

step: 20, loss: 0.581954836845398

step: 30, loss: 0.4348348081111908

step: 40, loss: 0.19119638204574585

step: 50, loss: 0.17578476667404175

step: 60, loss: 0.06256244331598282

step: 70, loss: 0.10499448329210281

step: 80, loss: 0.18002364039421082

step: 90, loss: 0.06543128937482834

step: 100, loss: 0.026006849482655525

step: 110, loss: 0.01140506099909544

step: 120, loss: 0.02085765264928341

step: 130, loss: 0.003016088856384158

```

step: 140, loss: 0.01600939780473709
step: 150, loss: 0.14576135575771332
step: 160, loss: 0.15281999111175537
step: 170, loss: 0.011519216001033783
step: 180, loss: 0.01258411630988121
step: 190, loss: 0.008226355537772179
step: 200, loss: 0.0023521804250776768
step: 210, loss: 0.0063596186228096485
step: 220, loss: 0.006256458815187216
step: 230, loss: 0.003521107602864504
/usr/local/lib/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=0.9898762654668166, f1=0.9887387387387387,
best_f1=0.9887387387387387
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007429099176079035
step: 10, loss: 0.0008102285210043192
step: 20, loss: 0.012365384958684444
step: 30, loss: 0.0019815973937511444
step: 40, loss: 0.020219292491674423
step: 50, loss: 0.004058694001287222
step: 60, loss: 0.003719736821949482
step: 70, loss: 0.0019142330856993794
step: 80, loss: 0.0011067750165238976
step: 90, loss: 0.0021149369422346354
step: 100, loss: 0.08540096133947372
step: 110, loss: 0.006728994194418192
step: 120, loss: 0.0004771519743371755
step: 130, loss: 0.020156938582658768
step: 140, loss: 0.169291153550148
step: 150, loss: 0.008120904676616192
step: 160, loss: 0.0073796785436570644
step: 170, loss: 0.01955031044781208
step: 180, loss: 0.0018238457851111889
step: 190, loss: 0.09494657069444656
step: 200, loss: 0.0038033947348594666
step: 210, loss: 0.01755194552242756
step: 220, loss: 0.0006110302056185901
step: 230, loss: 0.012065963819622993
/usr/local/lib/python3.7/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=0.9898534385569334, f1=0.9853107344632768,
best_f1=0.9887387387387387
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05017842724919319
step: 10, loss: 0.0015265482943505049
step: 20, loss: 0.002837514039129019
step: 30, loss: 0.0033845282159745693
step: 40, loss: 0.019981473684310913
step: 50, loss: 0.0006586327799595892
step: 60, loss: 0.0003149866533931345
step: 70, loss: 0.00037137148319743574
step: 80, loss: 0.007881652563810349
step: 90, loss: 0.00592432077974081
step: 100, loss: 0.001104481052607298
step: 110, loss: 0.04175984114408493
step: 120, loss: 0.0016925144009292126
step: 130, loss: 0.0012887591728940606
step: 140, loss: 0.005046315491199493
step: 150, loss: 0.001050058170221746
step: 160, loss: 0.0011237598955631256
step: 170, loss: 0.001911551458761096
step: 180, loss: 0.0017617419362068176
step: 190, loss: 0.0011051043402403593
step: 200, loss: 0.004002036526799202
step: 210, loss: 0.022470077499747276
step: 220, loss: 0.004613359458744526
step: 230, loss: 0.003551523433998227
/usr/local/lib/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=0.9921259842519685, f1=0.9865470852017937,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001506446860730648
step: 10, loss: 0.0005156049155630171
step: 20, loss: 0.00027524138567969203
step: 30, loss: 0.00028733216458931565
step: 40, loss: 0.0013499323977157474
step: 50, loss: 0.0077196513302624226

```

```

step: 60, loss: 0.0012038634158670902
step: 70, loss: 0.0017707409569993615
step: 80, loss: 0.06397733837366104
step: 90, loss: 0.011321756057441235
step: 100, loss: 0.001774149714037776
step: 110, loss: 0.004971191752701998
step: 120, loss: 0.006519681308418512
step: 130, loss: 0.02219279855489731
step: 140, loss: 0.00029203417943790555
step: 150, loss: 0.0009298103395849466
step: 160, loss: 0.0004500304057728499
step: 170, loss: 0.018001800402998924
step: 180, loss: 0.048561569303274155
step: 190, loss: 0.01224194560199976
step: 200, loss: 0.00036911264760419726
step: 210, loss: 0.00018779872334562242
step: 220, loss: 0.00032060645753517747
step: 230, loss: 0.00023597887775395066
/usr/local/lib/python3.7/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=0.9898989898989898, f1=0.987598647125141,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001641201670281589
step: 10, loss: 0.00028407108038663864
step: 20, loss: 0.00017253945406991988
step: 30, loss: 0.00010921114881057292
step: 40, loss: 0.00020280994067434222
step: 50, loss: 0.00013672994100488722
step: 60, loss: 0.0001618399692233652
step: 70, loss: 0.0001378279266646132
step: 80, loss: 0.0008516761008650064
step: 90, loss: 0.0024862890131771564
step: 100, loss: 0.0010464534861966968
step: 110, loss: 0.0012137920130044222
step: 120, loss: 0.027434099465608597
step: 130, loss: 0.029481982812285423
step: 140, loss: 0.00018880654533859342
step: 150, loss: 0.00017370648856740445
step: 160, loss: 0.01627037674188614
step: 170, loss: 0.08165264874696732
step: 180, loss: 0.0007293214439414442
step: 190, loss: 0.00017131899949163198
step: 200, loss: 0.0013484631199389696
step: 210, loss: 0.0004996404750272632

```



```

step: 220, loss: 0.00015543213521596044
step: 230, loss: 0.00022616340720560402
/usr/local/lib/python3.7/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=0.9887892376681614, f1=0.9887387387387387,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025161415338516235
step: 10, loss: 0.00040457327850162983
step: 20, loss: 0.0003316194051876664
step: 30, loss: 0.0002706901868805289
step: 40, loss: 0.0009094116976484656
step: 50, loss: 0.006495720241218805
step: 60, loss: 0.018920203670859337
step: 70, loss: 0.0007276175310835242
step: 80, loss: 0.00183706427924335
step: 90, loss: 0.00034378195414319634
step: 100, loss: 0.000842712470330298
step: 110, loss: 0.04260195046663284
step: 120, loss: 0.00019151323067490011
step: 130, loss: 0.001649007317610085
step: 140, loss: 0.00860647764056921
step: 150, loss: 0.00021460109564941376
step: 160, loss: 0.002758765360340476
step: 170, loss: 0.0002796807384584099
step: 180, loss: 0.00335004017688334
step: 190, loss: 0.002796160988509655
step: 200, loss: 0.0018560652388259768
step: 210, loss: 0.0005201722960919142
step: 220, loss: 0.0003519573074299842
step: 230, loss: 0.00015279435319826007
/usr/local/lib/python3.7/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=0.9876265466816648, f1=0.9887133182844244,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026243921369314194
step: 10, loss: 9.506350761512294e-05
step: 20, loss: 0.0003693308972287923
step: 30, loss: 0.00019712146604433656
step: 40, loss: 0.0006890140357427299
step: 50, loss: 9.562455670675263e-05

```

```

step: 60, loss: 0.0001762092870194465
step: 70, loss: 0.00011795339378295466
step: 80, loss: 8.007259748410434e-05
step: 90, loss: 0.0001689361233729869
step: 100, loss: 0.0006199849303811789
step: 110, loss: 0.05972600728273392
step: 120, loss: 0.137041836977005
step: 130, loss: 0.0026061036624014378
step: 140, loss: 0.0003373298095539212
step: 150, loss: 0.0011047348380088806
step: 160, loss: 0.0010267328470945358
step: 170, loss: 0.00014349287084769458
step: 180, loss: 0.0002685338840819895
step: 190, loss: 0.00017772313731256872
step: 200, loss: 0.0002663401246536523
step: 210, loss: 0.00026021775556728244
step: 220, loss: 0.0032280099112540483
step: 230, loss: 0.02614057995378971
/usr/local/lib/python3.7/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=0.9887387387387387, f1=0.9876265466816648,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0201959740370512
step: 10, loss: 0.023767471313476562
step: 20, loss: 0.0001450818672310561
step: 30, loss: 0.00012704241089522839
step: 40, loss: 0.00032061568344943225
step: 50, loss: 0.000710988650098443
step: 60, loss: 0.0002404776751063764
step: 70, loss: 0.00012853159569203854
step: 80, loss: 0.006744599901139736
step: 90, loss: 0.0009115033899433911
step: 100, loss: 0.00012155913282185793
step: 110, loss: 0.00016008537204470485
step: 120, loss: 0.00010958095663227141
step: 130, loss: 0.000549366872292012
step: 140, loss: 0.000882986350916326
step: 150, loss: 0.00042247213423252106
step: 160, loss: 0.00022730279306415468
step: 170, loss: 0.00030673545552417636
step: 180, loss: 0.00011901622201548889
step: 190, loss: 0.00020954936917405576
step: 200, loss: 0.0015033393865451217
step: 210, loss: 0.0012813060311600566

```

```

step: 220, loss: 0.00011130058555863798
step: 230, loss: 0.017240844666957855
/usr/local/lib/python3.7/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=0.9887133182844244, f1=0.9898305084745763,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002145009580999613
step: 10, loss: 0.00013774116814602166
step: 20, loss: 0.00019096194591838866
step: 30, loss: 0.00015795518993400037
step: 40, loss: 0.000148353457916528
step: 50, loss: 8.935906953411177e-05
step: 60, loss: 0.00011952193744946271
step: 70, loss: 0.00020619144197553396
step: 80, loss: 0.027006998658180237
step: 90, loss: 0.0015789545141160488
step: 100, loss: 0.00012477388372644782
step: 110, loss: 9.464607865083963e-05
step: 120, loss: 0.03219817206263542
step: 130, loss: 0.00014458750956691802
step: 140, loss: 9.108462108997628e-05
step: 150, loss: 7.055951573420316e-05
step: 160, loss: 0.00012370760669000447
step: 170, loss: 5.082951975055039e-05
step: 180, loss: 0.0002596391423139721
step: 190, loss: 5.1118935516569763e-05
step: 200, loss: 0.00026235688710585237
step: 210, loss: 0.00016293689259327948
step: 220, loss: 0.025578266009688377
step: 230, loss: 0.01359973382204771
/usr/local/lib/python3.7/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=0.9899216125419933, f1=0.9865771812080537,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001236942334799096
step: 10, loss: 8.679273742018268e-05
step: 20, loss: 7.831138645997271e-05
step: 30, loss: 9.584465442458168e-05
step: 40, loss: 4.292298763175495e-05
step: 50, loss: 0.00011625929619185627

```

```

step: 60, loss: 0.028084388002753258
step: 70, loss: 7.406398071907461e-05
step: 80, loss: 0.0011569381458684802
step: 90, loss: 6.625577225349844e-05
step: 100, loss: 7.036308670649305e-05
step: 110, loss: 8.13375081634149e-05
step: 120, loss: 0.017323162406682968
step: 130, loss: 5.58869396627415e-05
step: 140, loss: 0.02919268235564232
step: 150, loss: 9.694430627860129e-05
step: 160, loss: 9.645515820011497e-05
step: 170, loss: 0.00012910547957289964
step: 180, loss: 0.0004933342570438981
step: 190, loss: 7.462254143320024e-05
step: 200, loss: 9.873996168607846e-05
step: 210, loss: 5.706540832761675e-05
step: 220, loss: 0.00040720266406424344
step: 230, loss: 7.277901750057936e-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=0.9910313901345291, f1=0.9854423292273236,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.951751129236072e-05
step: 10, loss: 8.286422962555662e-05
step: 20, loss: 3.529938112478703e-05
step: 30, loss: 0.0006775879301130772
step: 40, loss: 0.003557015908882022
step: 50, loss: 0.002886663656681776
step: 60, loss: 6.476640555774793e-05
step: 70, loss: 6.943005428183824e-05
step: 80, loss: 9.966093057300895e-05
step: 90, loss: 5.153947131475434e-05
step: 100, loss: 0.00012871576473116875
step: 110, loss: 4.732716843136586e-05
step: 120, loss: 9.936459537129849e-05
step: 130, loss: 4.681557766161859e-05
step: 140, loss: 4.420911500346847e-05
step: 150, loss: 3.946381184505299e-05
step: 160, loss: 4.5035878429189324e-05
step: 170, loss: 0.017224809154868126
step: 180, loss: 0.00012756814248859882
step: 190, loss: 0.00018286111298948526
step: 200, loss: 9.912136738421395e-05
step: 210, loss: 4.975200135959312e-05

```

```

step: 220, loss: 5.6925338867586106e-05
step: 230, loss: 0.021190324798226357
/usr/local/lib/python3.7/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=0.9876543209876544, f1=0.9876265466816648,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.373409269144759e-05
step: 10, loss: 0.00010484583617653698
step: 20, loss: 4.1464274545433e-05
step: 30, loss: 0.00012417799734976143
step: 40, loss: 4.3017000280087814e-05
step: 50, loss: 0.00022231752518564463
step: 60, loss: 0.014994103461503983
step: 70, loss: 7.346548954956234e-05
step: 80, loss: 4.7467678086832166e-05
step: 90, loss: 7.711430953349918e-05
step: 100, loss: 5.4100266424939036e-05
step: 110, loss: 7.820573955541477e-05
step: 120, loss: 5.787964619230479e-05
step: 130, loss: 0.00015318150690291077
step: 140, loss: 0.0001353186380583793
step: 150, loss: 9.4414601335302e-05
step: 160, loss: 0.02923688292503357
step: 170, loss: 0.0013238402316346765
step: 180, loss: 6.160110206110403e-05
step: 190, loss: 0.0002148511412087828
step: 200, loss: 0.002541473601013422
step: 210, loss: 0.00011087075836258009
step: 220, loss: 0.0352785550057888
step: 230, loss: 6.980601028772071e-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=0.9899216125419933, f1=0.9865168539325843,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.99057272868231e-05
step: 10, loss: 0.0002826691197697073
step: 20, loss: 0.0001718949351925403
step: 30, loss: 0.024738572537899017
step: 40, loss: 0.00037739388062618673
step: 50, loss: 0.00011980416456935927

```

```

step: 60, loss: 7.938107592053711e-05
step: 70, loss: 7.423902570735663e-05
step: 80, loss: 0.0005386802367866039
step: 90, loss: 3.309757448732853e-05
step: 100, loss: 6.114591815276071e-05
step: 110, loss: 5.98789265495725e-05
step: 120, loss: 6.117128214100376e-05
step: 130, loss: 0.00011154075764352456
step: 140, loss: 6.75217597745359e-05
step: 150, loss: 0.016438789665699005
step: 160, loss: 5.100518319522962e-05
step: 170, loss: 6.053837205399759e-05
step: 180, loss: 0.00010713740630308166
step: 190, loss: 4.233788058627397e-05
step: 200, loss: 6.380190461641178e-05
step: 210, loss: 0.00016268157924059778
step: 220, loss: 4.181167605565861e-05
step: 230, loss: 4.754120163852349e-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=0.9899441340782122, f1=0.9844097995545658,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.757540879538283e-05
step: 10, loss: 2.8843056497862563e-05
step: 20, loss: 0.017149491235613823
step: 30, loss: 0.00011270265531493351
step: 40, loss: 3.8937097997404635e-05
step: 50, loss: 5.607601269730367e-05
step: 60, loss: 3.952956831199117e-05
step: 70, loss: 7.783593173371628e-05
step: 80, loss: 0.00024792016483843327
step: 90, loss: 0.00012041651643812656
step: 100, loss: 0.009642204269766808
step: 110, loss: 0.00012747228902298957
step: 120, loss: 4.292371522751637e-05
step: 130, loss: 0.00016471250273752958
step: 140, loss: 4.8178892029682174e-05
step: 150, loss: 6.766706064809114e-05
step: 160, loss: 3.425567410886288e-05
step: 170, loss: 3.7533351132879034e-05
step: 180, loss: 5.9618865634547547e-05
step: 190, loss: 0.007907143794000149
step: 200, loss: 5.146743933437392e-05
step: 210, loss: 0.015593930147588253

```

```

step: 220, loss: 7.644102879567072e-05
step: 230, loss: 2.2205944333109073e-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=0.9910313901345291, f1=0.9832026875699889,
best_f1=0.9865470852017937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.5447683330858126e-05
step: 10, loss: 5.1465034630382434e-05
step: 20, loss: 4.4723128667101264e-05
step: 30, loss: 6.586869858438149e-05
step: 40, loss: 7.210140756797045e-05
step: 50, loss: 0.00011389826977392659
step: 60, loss: 8.063896530074999e-05
step: 70, loss: 4.580676613841206e-05
step: 80, loss: 0.019705621525645256
step: 90, loss: 2.807275996019598e-05
step: 100, loss: 6.47683918941766e-05
step: 110, loss: 4.8387981223640963e-05
step: 120, loss: 6.286843563430011e-05
step: 130, loss: 8.792244625510648e-05
step: 140, loss: 0.16307465732097626
step: 150, loss: 0.0001028789920383133
step: 160, loss: 6.630246934946626e-05
step: 170, loss: 0.0003410857461858541
step: 180, loss: 4.480864299694076e-05
step: 190, loss: 0.01896965503692627
step: 200, loss: 4.346388959675096e-05
step: 210, loss: 0.026103435084223747
step: 220, loss: 0.00012098540901206434
step: 230, loss: 8.187899948097765e-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=0.9899216125419933, f1=0.9832026875699889,
best_f1=0.9865470852017937
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, 240.53it/s]
load_f1 = 0.9921259842519685
real_f1 = 0.9921259842519685
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, 259.68it/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: 0.7914928793907166
step: 10, loss: 0.4095258414745331
step: 20, loss: 0.47182488441467285
step: 30, loss: 0.3939504325389862
step: 40, loss: 0.2812733054161072
step: 50, loss: 0.1358129233121872
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.1635420173406601
step: 70, loss: 0.19460691511631012
step: 80, loss: 0.04668748006224632
step: 90, loss: 0.15860320627689362
step: 100, loss: 0.3873157203197479
step: 110, loss: 0.03473090007901192
step: 120, loss: 0.057167865335941315
step: 130, loss: 0.01545475609600544
step: 140, loss: 0.22983841598033905
step: 150, loss: 0.030087320134043694
step: 160, loss: 0.15547825396060944
step: 170, loss: 0.14518888294696808
step: 180, loss: 0.07396295666694641
step: 190, loss: 0.02472931146621704
step: 200, loss: 0.1168893352150917
step: 210, loss: 0.1348426192998886
step: 220, loss: 0.050577178597450256
step: 230, loss: 0.1041492447257042
step: 240, loss: 0.04903145506978035
step: 250, loss: 0.03633952513337135
step: 260, loss: 0.03411877900362015
step: 270, loss: 0.008716792799532413
step: 280, loss: 0.044486530125141144
step: 290, loss: 0.05572326481342316
step: 300, loss: 0.1313396394252777
step: 310, loss: 0.07721158117055893
step: 320, loss: 0.019433824345469475
step: 330, loss: 0.06565171480178833
step: 340, loss: 0.09408020973205566
step: 350, loss: 0.05348893254995346
step: 360, loss: 0.03319365158677101
step: 370, loss: 0.12026137858629227
step: 380, loss: 0.1312035620212555
step: 390, loss: 0.015814073383808136
step: 400, loss: 0.005513401702046394
step: 410, loss: 0.010989896021783352
step: 420, loss: 0.0036066295579075813
step: 430, loss: 0.009337985888123512
```

```

step: 440, loss: 0.07880006730556488
step: 450, loss: 0.013086648657917976
step: 460, loss: 0.2750997245311737
step: 470, loss: 0.14924301207065582
step: 480, loss: 0.2838696837425232
step: 490, loss: 0.040904074907302856
step: 500, loss: 0.006414215546101332
step: 510, loss: 0.0370950773358345
step: 520, loss: 0.04829533398151398
step: 530, loss: 0.12000174075365067
/usr/local/lib/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=0.9435370975268316, f1=0.9416705552963137,
best_f1=0.9416705552963137
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0692276805639267
step: 10, loss: 0.09232936054468155
step: 20, loss: 0.13034196197986603
step: 30, loss: 0.01684964820742607
step: 40, loss: 0.004147414118051529
step: 50, loss: 0.027554256841540337
step: 60, loss: 0.08497612178325653
step: 70, loss: 0.14492064714431763
step: 80, loss: 0.008187743835151196
step: 90, loss: 0.0034687987063080072
step: 100, loss: 0.29665765166282654
step: 110, loss: 0.021880043670535088
step: 120, loss: 0.0690760537981987
step: 130, loss: 0.025907760486006737
step: 140, loss: 0.04477793723344803
step: 150, loss: 0.06994164735078812
step: 160, loss: 0.009602304548025131
step: 170, loss: 0.08019793033599854
step: 180, loss: 0.001559709431603551
step: 190, loss: 0.004056269768625498
step: 200, loss: 0.008824813179671764
step: 210, loss: 0.002079491037875414
step: 220, loss: 0.09125089645385742
step: 230, loss: 0.014169988222420216
step: 240, loss: 0.1650925725698471
step: 250, loss: 0.03655434772372246

```

```

step: 260, loss: 0.01566281169652939
step: 270, loss: 0.10779132694005966
step: 280, loss: 0.0971878245472908
step: 290, loss: 0.05697408691048622
step: 300, loss: 0.014709815382957458
step: 310, loss: 0.022239001467823982
step: 320, loss: 0.015767332166433334
step: 330, loss: 0.015348712913691998
step: 340, loss: 0.027928601950407028
step: 350, loss: 0.030448298901319504
step: 360, loss: 0.025235222652554512
step: 370, loss: 0.00801867712289095
step: 380, loss: 0.055101022124290466
step: 390, loss: 0.021592913195490837
step: 400, loss: 0.034115858376026154
step: 410, loss: 0.0002582728920970112
step: 420, loss: 0.020914915949106216
step: 430, loss: 0.03512758016586304
step: 440, loss: 0.01249210350215435
step: 450, loss: 0.018424058333039284
step: 460, loss: 0.2032170295715332
step: 470, loss: 0.036759939044713974
step: 480, loss: 0.16676722466945648
step: 490, loss: 0.04738805815577507
step: 500, loss: 0.018345395103096962
step: 510, loss: 0.026606164872646332
step: 520, loss: 0.07871683686971664
step: 530, loss: 0.06921570003032684
/usr/local/lib/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=0.9498607242339833, f1=0.9453197405004634,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002794966334477067
step: 10, loss: 0.010506154038012028
step: 20, loss: 0.1696072667837143
step: 30, loss: 0.2085355520248413
step: 40, loss: 0.00494926143437624
step: 50, loss: 0.006424613296985626
step: 60, loss: 0.0023608170449733734
step: 70, loss: 0.02857743576169014

```

```
step: 80, loss: 0.0045630233362317085
step: 90, loss: 0.002966122003272176
step: 100, loss: 0.13299302756786346
step: 110, loss: 0.018403230234980583
step: 120, loss: 0.027512019500136375
step: 130, loss: 0.02424258179962635
step: 140, loss: 0.05721944198012352
step: 150, loss: 0.009584055282175541
step: 160, loss: 0.010153698734939098
step: 170, loss: 0.0035240475554019213
step: 180, loss: 0.0022355751134455204
step: 190, loss: 0.003793273353949189
step: 200, loss: 0.029490778222680092
step: 210, loss: 0.03432921692728996
step: 220, loss: 0.0065483455546200275
step: 230, loss: 0.03662756830453873
step: 240, loss: 0.007355409674346447
step: 250, loss: 0.005932093597948551
step: 260, loss: 0.0006136429728940129
step: 270, loss: 0.0009205489768646657
step: 280, loss: 0.005356735549867153
step: 290, loss: 0.04227026551961899
step: 300, loss: 0.008423153311014175
step: 310, loss: 0.12410252541303635
step: 320, loss: 0.09372462332248688
step: 330, loss: 0.003642335766926408
step: 340, loss: 0.0007946657715365291
step: 350, loss: 0.010510646738111973
step: 360, loss: 0.008235319517552853
step: 370, loss: 0.0015423217555508018
step: 380, loss: 0.001454210956580937
step: 390, loss: 0.0041845799423754215
step: 400, loss: 0.006421077996492386
step: 410, loss: 0.006367722991853952
step: 420, loss: 0.022502722218632698
step: 430, loss: 0.003780932165682316
step: 440, loss: 0.025776734575629234
step: 450, loss: 0.0009940393501892686
step: 460, loss: 0.09192444384098053
step: 470, loss: 0.0035617637913674116
step: 480, loss: 0.14113549888134003
step: 490, loss: 0.03624119237065315
step: 500, loss: 0.033676888793706894
step: 510, loss: 0.007041318342089653
step: 520, loss: 0.08834034949541092
step: 530, loss: 0.036554452031850815
/usr/local/lib/python3.7/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=0.9463548830811555, f1=0.9413936317489617,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0037688964512199163
step: 10, loss: 0.005966775584965944
step: 20, loss: 0.03544808551669121
step: 30, loss: 0.01419342402368784
step: 40, loss: 0.01068107970058918
step: 50, loss: 0.009379249066114426
step: 60, loss: 0.0002723316429182887
step: 70, loss: 0.0010575001360848546
step: 80, loss: 0.04153802990913391
step: 90, loss: 0.004889694973826408
step: 100, loss: 0.026521092280745506
step: 110, loss: 0.0011214715195819736
step: 120, loss: 0.0002901942061726004
step: 130, loss: 0.001326658413745463
step: 140, loss: 0.0009368943283334374
step: 150, loss: 0.0003210847789887339
step: 160, loss: 0.05441009998321533
step: 170, loss: 0.0010699373669922352
step: 180, loss: 0.003852773690596223
step: 190, loss: 0.0007716575055383146
step: 200, loss: 0.0006568230455741286
step: 210, loss: 0.0030750290025025606
step: 220, loss: 0.0017533765640109777
step: 230, loss: 0.25323420763015747
step: 240, loss: 0.0014431675663217902
step: 250, loss: 0.004202273674309254
step: 260, loss: 0.10826436430215836
step: 270, loss: 0.00865914486348629
step: 280, loss: 0.0011900286190211773
step: 290, loss: 0.0021932676900178194
step: 300, loss: 0.015622529201209545
step: 310, loss: 0.0016564094694331288
step: 320, loss: 0.014816305600106716
step: 330, loss: 0.0015847699251025915
step: 340, loss: 0.06203102320432663
step: 350, loss: 0.00017460160597693175
step: 360, loss: 0.001988765085116029
step: 370, loss: 0.13592056930065155
step: 380, loss: 0.01640760712325573
step: 390, loss: 0.016234250739216805
step: 400, loss: 0.015124019235372543
step: 410, loss: 0.01644458808004856
```

```
step: 420, loss: 0.06901621073484421
step: 430, loss: 0.00891869142651558
step: 440, loss: 0.07784543931484222
step: 450, loss: 0.015458531677722931
step: 460, loss: 0.007806996814906597
step: 470, loss: 0.011725440621376038
step: 480, loss: 0.00605213874951005
step: 490, loss: 0.009452804923057556
step: 500, loss: 0.004064507316797972
step: 510, loss: 0.0029767232481390238
step: 520, loss: 0.01629229076206684
step: 530, loss: 0.013001538813114166
/usr/local/lib/python3.7/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=0.9456029011786038, f1=0.9410150891632373,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0030626910738646984
step: 10, loss: 0.0006226266268640757
step: 20, loss: 0.0065231299959123135
step: 30, loss: 0.008114853873848915
step: 40, loss: 0.02675708942115307
step: 50, loss: 0.016431203112006187
step: 60, loss: 0.0011901231482625008
step: 70, loss: 0.0002718810283113271
step: 80, loss: 0.0018320102244615555
step: 90, loss: 0.04413676634430885
step: 100, loss: 0.0008055656799115241
step: 110, loss: 0.03536228835582733
step: 120, loss: 0.04435649886727333
step: 130, loss: 0.0030392445623874664
step: 140, loss: 0.0011961538111791015
step: 150, loss: 0.0016268615145236254
step: 160, loss: 0.01766548678278923
step: 170, loss: 0.00153254228644073
step: 180, loss: 0.003946834709495306
step: 190, loss: 0.0001096860141842626
step: 200, loss: 0.0003199981292709708
step: 210, loss: 0.0002826155978254974
step: 220, loss: 0.00010912831203313544
step: 230, loss: 0.0015179410111159086
step: 240, loss: 0.0027236007153987885
step: 250, loss: 0.0003924874763470143
step: 260, loss: 0.02384301833808422
step: 270, loss: 0.020446304231882095
```

```

step: 280, loss: 0.010865229181945324
step: 290, loss: 0.00044073472963646054
step: 300, loss: 0.00085788982687518
step: 310, loss: 0.0001482652296544984
step: 320, loss: 0.009380168281495571
step: 330, loss: 0.001704991445876658
step: 340, loss: 0.0031137624755501747
step: 350, loss: 0.05295650288462639
step: 360, loss: 0.027968158945441246
step: 370, loss: 0.010750864632427692
step: 380, loss: 0.022825071588158607
step: 390, loss: 0.006102582905441523
step: 400, loss: 0.001468016649596393
step: 410, loss: 0.001363749266602099
step: 420, loss: 0.028040887787938118
step: 430, loss: 0.007486057002097368
step: 440, loss: 0.004943004809319973
step: 450, loss: 0.005237923003733158
step: 460, loss: 0.0007864272338338196
step: 470, loss: 0.00037680970854125917
step: 480, loss: 0.0010372489923611283
step: 490, loss: 0.019307715818285942
step: 500, loss: 0.0010122049134224653
step: 510, loss: 0.0025319992564618587
step: 520, loss: 0.00021604359790217131
step: 530, loss: 0.01913081295788288
/usr/local/lib/python3.7/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=0.9456928838951311, f1=0.943609022556391,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015486404299736023
step: 10, loss: 0.013414083980023861
step: 20, loss: 0.0002685984945856035
step: 30, loss: 0.09539337456226349
step: 40, loss: 0.00048547046026214957
step: 50, loss: 0.07005362957715988
step: 60, loss: 0.0028858568985015154
step: 70, loss: 0.03939613699913025
step: 80, loss: 0.00021952039969619364
step: 90, loss: 0.005129201803356409
step: 100, loss: 0.031086627393960953
step: 110, loss: 0.02177583798766136
step: 120, loss: 0.00033947662450373173
step: 130, loss: 0.0002852371835615486

```



```

step: 140, loss: 0.0014165283646434546
step: 150, loss: 0.05313081294298172
step: 160, loss: 0.11789858341217041
step: 170, loss: 0.00017823695088736713
step: 180, loss: 0.0003222897066734731
step: 190, loss: 0.001466164132580161
step: 200, loss: 5.7525470765540376e-05
step: 210, loss: 0.0045832740142941475
step: 220, loss: 0.0002896032528951764
step: 230, loss: 0.00015739262744318694
step: 240, loss: 6.932709948159754e-05
step: 250, loss: 0.0023035211488604546
step: 260, loss: 0.0001626685116207227
step: 270, loss: 4.422521305968985e-05
step: 280, loss: 0.006208225153386593
step: 290, loss: 0.00026313410489819944
step: 300, loss: 0.00014033084153197706
step: 310, loss: 0.00022021667973604053
step: 320, loss: 0.05437209829688072
step: 330, loss: 0.016912922263145447
step: 340, loss: 0.0020182100124657154
step: 350, loss: 0.03337232023477554
step: 360, loss: 0.00546937994658947
step: 370, loss: 0.0001506218541180715
step: 380, loss: 0.009145436808466911
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 390, loss: 0.000941262929700315
step: 400, loss: 4.754027395392768e-05
step: 410, loss: 4.997995347366668e-05
step: 420, loss: 0.00030555683770217
step: 430, loss: 5.678412344423123e-05
step: 440, loss: 0.0004120935336686671
step: 450, loss: 0.0008950492483563721
step: 460, loss: 0.0019516960019245744
step: 470, loss: 0.01359502412378788
step: 480, loss: 0.006151104345917702
step: 490, loss: 7.973376341396943e-05
step: 500, loss: 0.022106701508164406
step: 510, loss: 0.00011025607091141865
step: 520, loss: 0.001930435188114643
step: 530, loss: 0.0017638277495279908
/usr/local/lib/python3.7/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=0.9475164011246484, f1=0.9442622950819672,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005418607499450445
step: 10, loss: 0.020996594801545143
step: 20, loss: 0.0020008746068924665
step: 30, loss: 0.01450619287788868
step: 40, loss: 0.019176766276359558
step: 50, loss: 0.00019872926350217313
step: 60, loss: 0.001404489972628653
step: 70, loss: 0.0015124290948733687
step: 80, loss: 0.0022625159472227097
step: 90, loss: 7.940366049297154e-05
step: 100, loss: 0.00014046243450138718
step: 110, loss: 0.0001236082025570795
step: 120, loss: 7.282933074748144e-05
step: 130, loss: 0.00030415423680096865
step: 140, loss: 0.0008519915281794965
step: 150, loss: 7.051568536553532e-05
step: 160, loss: 0.0007458277978003025
step: 170, loss: 0.00022894424910191447
step: 180, loss: 5.1885959692299366e-05
step: 190, loss: 0.0006266737473197281
step: 200, loss: 5.1562041335273534e-05
step: 210, loss: 0.002421507379040122
step: 220, loss: 6.904741167090833e-05
step: 230, loss: 0.00036590578383766115
step: 240, loss: 0.003232378512620926
step: 250, loss: 0.0005617181304842234
step: 260, loss: 0.005115207750350237
step: 270, loss: 0.0002837692736648023
step: 280, loss: 0.01417718268930912
step: 290, loss: 0.0012690238654613495
step: 300, loss: 0.00016333341773133725
step: 310, loss: 0.0004005449591204524
step: 320, loss: 0.09283675253391266
step: 330, loss: 0.002646669512614608
step: 340, loss: 0.012438083998858929
step: 350, loss: 0.0536639466881752
step: 360, loss: 0.03941522538661957
step: 370, loss: 0.00025167103740386665
step: 380, loss: 0.0016188295558094978
step: 390, loss: 4.822441769647412e-05
step: 400, loss: 2.9771810659440234e-05
step: 410, loss: 0.00132499891333282
step: 420, loss: 0.0001335895067313686
step: 430, loss: 4.7407535021193326e-05
step: 440, loss: 0.00022622114920523018
step: 450, loss: 0.0004464068333618343
step: 460, loss: 0.001867462182417512
```

```

step: 470, loss: 0.0029596828389912844
step: 480, loss: 0.0016709393821656704
step: 490, loss: 0.0002057024248642847
step: 500, loss: 4.869055555900559e-05
step: 510, loss: 0.0008393017924390733
step: 520, loss: 0.0001506760308984667
step: 530, loss: 7.090438884915784e-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=0.9436619718309859, f1=0.9371482176360225,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006699653808027506
step: 10, loss: 0.0008931848569773138
step: 20, loss: 0.013278587721288204
step: 30, loss: 0.013986179605126381
step: 40, loss: 0.00014410613221116364
step: 50, loss: 0.00031862634932622313
step: 60, loss: 0.07037831097841263
step: 70, loss: 0.0007415346917696297
step: 80, loss: 0.00025897088926285505
step: 90, loss: 0.0002870905736926943
step: 100, loss: 0.005722233094274998
step: 110, loss: 0.0007293516537174582
step: 120, loss: 0.005928094033151865
step: 130, loss: 0.0021343177650123835
step: 140, loss: 3.068067962885834e-05
step: 150, loss: 0.0006385350716300309
step: 160, loss: 0.0001835734728956595
step: 170, loss: 0.0003854862879961729
step: 180, loss: 0.00011538221588125452
step: 190, loss: 0.00014754149015061557
step: 200, loss: 0.00024208588001783937
step: 210, loss: 0.007110758684575558
step: 220, loss: 0.0005500481929630041
step: 230, loss: 6.840760033810511e-05
step: 240, loss: 0.0006825583404861391
step: 250, loss: 0.0025285258889198303
step: 260, loss: 0.00927550345659256
step: 270, loss: 4.562524554785341e-05
step: 280, loss: 0.0005509195616468787
step: 290, loss: 0.03572872653603554
step: 300, loss: 0.00040351386996917427
step: 310, loss: 0.02371562086045742
step: 320, loss: 0.0010558698559179902

```

```

step: 330, loss: 0.010111489333212376
step: 340, loss: 0.001812006812542677
step: 350, loss: 0.032767556607723236
step: 360, loss: 0.004976773168891668
step: 370, loss: 0.0016249502077698708
step: 380, loss: 0.00020209686772432178
step: 390, loss: 0.0004237940302118659
step: 400, loss: 0.012441528029739857
step: 410, loss: 6.679050420643762e-05
step: 420, loss: 6.855746323708445e-05
step: 430, loss: 0.0011604566825553775
step: 440, loss: 0.00035242902231402695
step: 450, loss: 0.002141656121239066
step: 460, loss: 0.0045857359655201435
step: 470, loss: 0.0002559037529863417
step: 480, loss: 0.0005889332387596369
step: 490, loss: 0.00046675303019583225
step: 500, loss: 0.010684510692954063
step: 510, loss: 0.00016597742796875536
step: 520, loss: 0.00038943407707847655
step: 530, loss: 0.0003732991754077375
/usr/local/lib/python3.7/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=0.9478584729981377, f1=0.9436422915696321,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016501997597515583
step: 10, loss: 9.505341586191207e-05
step: 20, loss: 0.00015644407540094107
step: 30, loss: 0.003233987372368574
step: 40, loss: 0.0003159975167363882
step: 50, loss: 0.00020014964684378356
step: 60, loss: 0.0002525441232137382
step: 70, loss: 0.12547904253005981
step: 80, loss: 0.0008369461284019053
step: 90, loss: 0.02382240816950798
step: 100, loss: 0.009718927554786205
step: 110, loss: 0.0005362095544114709
step: 120, loss: 3.24799693771638e-05
step: 130, loss: 0.0004324206383898854
step: 140, loss: 0.0004608290910255164
step: 150, loss: 6.419533747248352e-05
step: 160, loss: 4.8115882236743346e-05
step: 170, loss: 0.0001177022495539859
step: 180, loss: 5.309662810759619e-05

```

```

step: 190, loss: 0.0001143209810834378
step: 200, loss: 0.0005678486777469516
step: 210, loss: 2.5670424292911775e-05
step: 220, loss: 0.001596866874024272
step: 230, loss: 2.5715038646012545e-05
step: 240, loss: 0.00012250998406670988
step: 250, loss: 0.00023395173775497824
step: 260, loss: 4.736506525659934e-05
step: 270, loss: 0.0006963419727981091
step: 280, loss: 3.999570253654383e-05
step: 290, loss: 2.453037268423941e-05
step: 300, loss: 7.274741074070334e-05
step: 310, loss: 3.0009347028681077e-05
step: 320, loss: 0.00024664521333761513
step: 330, loss: 4.166825965512544e-05
step: 340, loss: 0.0004788602236658335
step: 350, loss: 3.764397115446627e-05
step: 360, loss: 0.029853085055947304
step: 370, loss: 5.764067464042455e-05
step: 380, loss: 2.459021379763726e-05
step: 390, loss: 0.0004043321532662958
step: 400, loss: 0.0028634690679609776
step: 410, loss: 8.214142872020602e-05
step: 420, loss: 0.0014690053649246693
step: 430, loss: 2.6758065359899774e-05
step: 440, loss: 0.0014435534831136465
step: 450, loss: 3.6269815609557554e-05
step: 460, loss: 6.724326522089541e-05
step: 470, loss: 2.140886317647528e-05
step: 480, loss: 6.428256165236235e-05
step: 490, loss: 6.658305937889963e-05
step: 500, loss: 8.168452768586576e-05
step: 510, loss: 0.0001239276898559183
step: 520, loss: 0.0001140311433118768
step: 530, loss: 3.5473378375172615e-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=0.9486940298507462, f1=0.9420560747663551,
best_f1=0.9453197405004634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010463051876286045
step: 10, loss: 1.5158069800236262e-05
step: 20, loss: 3.23642925650347e-05
step: 30, loss: 1.8804987121257e-05
step: 40, loss: 3.045618359465152e-05

```

step: 50, loss: 2.020552165049594e-05
step: 60, loss: 4.1860505007207394e-05
step: 70, loss: 0.0014920111279934645
step: 80, loss: 0.0005409959703683853
step: 90, loss: 5.788561611552723e-05
step: 100, loss: 2.708596184675116e-05
step: 110, loss: 0.0005601060693152249
step: 120, loss: 3.277020368841477e-05
step: 130, loss: 2.226553078799043e-05
step: 140, loss: 4.5238324673846364e-05
step: 150, loss: 2.6474110200069845e-05
step: 160, loss: 0.00039233299321494997
step: 170, loss: 5.18385240866337e-05
step: 180, loss: 0.08356323093175888
step: 190, loss: 5.249312744126655e-05
step: 200, loss: 0.0171889029443264
step: 210, loss: 0.00014987392933107913
step: 220, loss: 0.002572550205513835
step: 230, loss: 0.024051513522863388
step: 240, loss: 0.006262191571295261
step: 250, loss: 8.71413285494782e-05
step: 260, loss: 0.01015500444918871
step: 270, loss: 4.585980423144065e-05
step: 280, loss: 0.004625749308615923
step: 290, loss: 0.0003880732401739806
step: 300, loss: 7.04617632436566e-05
step: 310, loss: 3.025196201633662e-05
step: 320, loss: 0.0011332198046147823
step: 330, loss: 0.00014546063903253525
step: 340, loss: 4.4011729187332094e-05
step: 350, loss: 9.309041342930868e-05
step: 360, loss: 0.00018802013073582202
step: 370, loss: 0.0013826512731611729
step: 380, loss: 3.413370723137632e-05
step: 390, loss: 3.011802982655354e-05
step: 400, loss: 2.309262526978273e-05
step: 410, loss: 2.655307798704598e-05
step: 420, loss: 0.07397953420877457
step: 430, loss: 2.5614312107791193e-05
step: 440, loss: 4.5465923903975636e-05
step: 450, loss: 0.020655790343880653
step: 460, loss: 7.623346755281091e-05
step: 470, loss: 2.1624875444103964e-05
step: 480, loss: 3.303063931525685e-05
step: 490, loss: 0.04953542351722717
step: 500, loss: 9.234401659341529e-05
step: 510, loss: 0.000560711428988725
step: 520, loss: 5.06672658957541e-05

```

step: 530, loss: 4.9173580919159576e-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=0.9501630181648812, f1=0.9463955637707948,
best_f1=0.9463955637707948
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029186302796006203
step: 10, loss: 0.0015913101378828287
step: 20, loss: 2.6719464585767128e-05
step: 30, loss: 5.596087794401683e-05
step: 40, loss: 2.47018597292481e-05
step: 50, loss: 0.0006883436581119895
step: 60, loss: 7.156623905757442e-05
step: 70, loss: 5.95708770561032e-05
step: 80, loss: 0.00015840862761251628
step: 90, loss: 0.0016372257377952337
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 0.0008508110186085105
step: 110, loss: 4.294044993002899e-05
step: 120, loss: 0.10332990437746048
step: 130, loss: 5.5527081713080406e-05
step: 140, loss: 0.0012518969597294927
step: 150, loss: 7.897245814092457e-05
step: 160, loss: 6.613470031879842e-05
step: 170, loss: 0.0004878908512182534
step: 180, loss: 4.8582158342469484e-05
step: 190, loss: 0.0009826472960412502
step: 200, loss: 0.0011242787586525083
step: 210, loss: 7.124804687919095e-05
step: 220, loss: 3.1991457944968715e-05
step: 230, loss: 2.9327693482628092e-05
step: 240, loss: 5.8417375839781016e-05
step: 250, loss: 4.440058546606451e-05
step: 260, loss: 0.00023973127827048302
step: 270, loss: 0.001082144328393042
step: 280, loss: 0.00022550096036866307
step: 290, loss: 0.02196669951081276
step: 300, loss: 0.002950799185782671
step: 310, loss: 0.005372829269617796
step: 320, loss: 0.021235540509223938
step: 330, loss: 0.001349150319583714

```

```

step: 340, loss: 0.0010805710917338729
step: 350, loss: 0.0013090325519442558
step: 360, loss: 4.9776579544413835e-05
step: 370, loss: 1.7009453586069867e-05
step: 380, loss: 2.1080979422549717e-05
step: 390, loss: 0.044485095888376236
step: 400, loss: 6.349867180688307e-05
step: 410, loss: 0.0008483376586809754
step: 420, loss: 0.0007640086696483195
step: 430, loss: 0.00034688637242652476
step: 440, loss: 9.871520887827501e-05
step: 450, loss: 3.3435586374253035e-05
step: 460, loss: 0.0019468954997137189
step: 470, loss: 0.009677628986537457
step: 480, loss: 0.00022105331299826503
step: 490, loss: 0.12508000433444977
step: 500, loss: 5.615127884084359e-05
step: 510, loss: 4.107719723833725e-05
step: 520, loss: 2.025027060881257e-05
step: 530, loss: 2.490606675564777e-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=0.9501165501165502, f1=0.9461966604823748,
best_f1=0.9463955637707948
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.510822327574715e-05
step: 10, loss: 0.0001868793769972399
step: 20, loss: 1.550450178910978e-05
step: 30, loss: 8.696938311913982e-05
step: 40, loss: 6.651130388490856e-05
step: 50, loss: 0.0030396808870136738
step: 60, loss: 0.023449571803212166
step: 70, loss: 0.00017649735673330724
step: 80, loss: 0.0013030925765633583
step: 90, loss: 5.260048419586383e-05
step: 100, loss: 0.0001344513875665143
step: 110, loss: 0.007907706312835217
step: 120, loss: 0.0660690888762474
step: 130, loss: 5.325419260771014e-05
step: 140, loss: 4.339453516877256e-05
step: 150, loss: 0.006152020301669836
step: 160, loss: 4.5107030018698424e-05
step: 170, loss: 0.002504161559045315
step: 180, loss: 0.00011436720524216071
step: 190, loss: 3.083519550273195e-05

```



```

step: 200, loss: 9.876995318336412e-05
step: 210, loss: 0.00010580885282251984
step: 220, loss: 3.442843444645405e-05
step: 230, loss: 0.002066673245280981
step: 240, loss: 0.00017768416728358716
step: 250, loss: 4.0342583815800026e-05
step: 260, loss: 0.00018416323291603476
step: 270, loss: 3.071302126045339e-05
step: 280, loss: 3.7201523809926584e-05
step: 290, loss: 0.009019712917506695
step: 300, loss: 0.00026031164452433586
step: 310, loss: 3.414372986298986e-05
step: 320, loss: 0.00010178402590099722
step: 330, loss: 5.990520367049612e-05
step: 340, loss: 0.0001274993846891448
step: 350, loss: 0.027462109923362732
step: 360, loss: 0.000622218765784055
step: 370, loss: 1.6294219676638022e-05
step: 380, loss: 2.293965735589154e-05
step: 390, loss: 3.152039789711125e-05
step: 400, loss: 1.629792677704245e-05
step: 410, loss: 0.00034717359812930226
step: 420, loss: 0.0008318793843500316
step: 430, loss: 0.00010745152394520119
step: 440, loss: 0.0011261694598942995
step: 450, loss: 0.0004907454131171107
step: 460, loss: 0.0006756147486157715
step: 470, loss: 1.7206815755343996e-05
step: 480, loss: 5.4635016567772254e-05
step: 490, loss: 0.007956656627357006
step: 500, loss: 0.004558156244456768
step: 510, loss: 2.5241963157895952e-05
step: 520, loss: 0.018320444971323013
step: 530, loss: 0.00037690327735617757
/usr/local/lib/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=0.9502093997208004, f1=0.9460966542750929,
best_f1=0.9460966542750929
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.935970643302426e-05
step: 10, loss: 4.8999245336744934e-05

```

step: 20, loss: 0.025336945429444313
step: 30, loss: 4.630510011338629e-05
step: 40, loss: 2.6933066692436114e-05
step: 50, loss: 2.2515010641654953e-05
step: 60, loss: 3.454784382483922e-05
step: 70, loss: 1.6797064745333046e-05
step: 80, loss: 3.1370924261864275e-05
step: 90, loss: 0.0004458119801711291
step: 100, loss: 0.0001123179608839564
step: 110, loss: 4.967287532053888e-05
step: 120, loss: 4.934422031510621e-05
step: 130, loss: 0.0010340515291318297
step: 140, loss: 0.0002122732694260776
step: 150, loss: 3.287198705947958e-05
step: 160, loss: 2.0678537111962214e-05
step: 170, loss: 3.962882328778505e-05
step: 180, loss: 3.2178806577576324e-05
step: 190, loss: 6.288247823249549e-05
step: 200, loss: 0.0017994260415434837
step: 210, loss: 0.0014703597407788038
step: 220, loss: 2.135257818736136e-05
step: 230, loss: 2.456633410474751e-05
step: 240, loss: 2.1058551283203997e-05
step: 250, loss: 0.041721753776073456
step: 260, loss: 2.917397978308145e-05
step: 270, loss: 0.00015197791799437255
step: 280, loss: 1.6308984413626604e-05
step: 290, loss: 0.002855968428775668
step: 300, loss: 2.1099647710798308e-05
step: 310, loss: 0.028231261298060417
step: 320, loss: 2.5930583433364518e-05
step: 330, loss: 0.0006761762197129428
step: 340, loss: 2.122255136782769e-05
step: 350, loss: 0.008360633626580238
step: 360, loss: 3.0718052585143596e-05
step: 370, loss: 6.407951877918094e-05
step: 380, loss: 0.000601640495005995
step: 390, loss: 0.2417214959859848
step: 400, loss: 1.927790253830608e-05
step: 410, loss: 4.387473381939344e-05
step: 420, loss: 3.399317211005837e-05
step: 430, loss: 0.0008858813671395183
step: 440, loss: 0.00012500806769821793
step: 450, loss: 0.00014696696598548442
step: 460, loss: 3.526498403516598e-05
step: 470, loss: 0.026783453300595284
step: 480, loss: 0.0010371390962973237
step: 490, loss: 4.8545178287895396e-05

```

step: 500, loss: 3.2635845855111256e-05
step: 510, loss: 3.420573193579912e-05
step: 520, loss: 1.6808211512397975e-05
step: 530, loss: 4.277041807654314e-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 13: dev_f1=0.9503019043195542, f1=0.9447795823665893,
best_f1=0.9447795823665893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00044114506454207003
step: 10, loss: 3.756271325983107e-05
step: 20, loss: 2.554160528234206e-05
step: 30, loss: 6.790002225898206e-05
step: 40, loss: 6.389962072717026e-05
step: 50, loss: 5.497123856912367e-05
step: 60, loss: 6.121933984104544e-05
step: 70, loss: 1.8037419067695737e-05
step: 80, loss: 0.00011941391858272254
step: 90, loss: 1.4066476069274358e-05
step: 100, loss: 6.527030927827582e-05
step: 110, loss: 2.3869994038250297e-05
step: 120, loss: 6.66190026095137e-05
step: 130, loss: 0.00013185219722799957
step: 140, loss: 0.0005622272146865726
step: 150, loss: 0.00011951506894547492
step: 160, loss: 0.0011981657007709146
step: 170, loss: 0.00022217836522031575
step: 180, loss: 0.00023891137971077114
step: 190, loss: 9.177206084132195e-05
step: 200, loss: 3.172966535203159e-05
step: 210, loss: 0.00010938010382233188
step: 220, loss: 9.577556193107739e-05
step: 230, loss: 0.0015435044188052416
step: 240, loss: 1.7117461538873613e-05
step: 250, loss: 3.1145398679655045e-05
step: 260, loss: 1.1894741874129977e-05
step: 270, loss: 0.0033100333530455828
step: 280, loss: 2.163173303415533e-05
step: 290, loss: 0.0019347796915099025
step: 300, loss: 2.6916932256426662e-05
step: 310, loss: 2.0328099708422087e-05

```

```

step: 320, loss: 3.359191396157257e-05
step: 330, loss: 0.0005562199512496591
step: 340, loss: 4.334151526563801e-05
step: 350, loss: 0.0056370398961007595
step: 360, loss: 0.001898852875456214
step: 370, loss: 0.006633399520069361
step: 380, loss: 7.335875125136226e-05
step: 390, loss: 0.00027202890487387776
step: 400, loss: 1.7937056327355094e-05
step: 410, loss: 2.416851930320263e-05
step: 420, loss: 3.9155165723059326e-05
step: 430, loss: 2.054829201370012e-05
step: 440, loss: 1.2289635378692765e-05
step: 450, loss: 1.5262332453858107e-05
step: 460, loss: 6.45170221105218e-05
step: 470, loss: 8.239733870141208e-05
step: 480, loss: 2.9727538276347332e-05
step: 490, loss: 1.9385817722650245e-05
step: 500, loss: 2.028751805482898e-05
step: 510, loss: 2.6496854843571782e-05
step: 520, loss: 3.2584910513833165e-05
step: 530, loss: 2.0913095795549452e-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=0.9486348912540491, f1=0.9455216989843028,
best_f1=0.9447795823665893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 1.9225417418056168e-05
step: 10, loss: 0.00014056525833439082
step: 20, loss: 4.649149559554644e-05
step: 30, loss: 6.095104981795885e-05
step: 40, loss: 2.3258971850736998e-05
step: 50, loss: 0.0005721314810216427
step: 60, loss: 1.2878173038188834e-05
step: 70, loss: 0.00046854838728904724
step: 80, loss: 2.9200715289334767e-05
step: 90, loss: 1.710613287286833e-05
step: 100, loss: 2.9437362172757275e-05
step: 110, loss: 0.013197137042880058
step: 120, loss: 0.0014619810972362757
step: 130, loss: 0.0004493217566050589
step: 140, loss: 1.2300803064135835e-05
step: 150, loss: 0.00022927398094907403
step: 160, loss: 3.045179801119957e-05
step: 170, loss: 8.0150603025686e-05

```

```

step: 180, loss: 1.6115367543534376e-05
step: 190, loss: 6.759171083103865e-05
step: 200, loss: 1.9047065507038496e-05
step: 210, loss: 2.042148116743192e-05
step: 220, loss: 1.210337632073788e-05
step: 230, loss: 1.3634394235850777e-05
step: 240, loss: 2.5639636078267358e-05
step: 250, loss: 8.882214751793072e-05
step: 260, loss: 9.276803757529706e-05
step: 270, loss: 0.0003883053141180426
step: 280, loss: 0.00045638991286978126
step: 290, loss: 0.0011343752266839147
step: 300, loss: 0.004388665314763784
step: 310, loss: 5.7069857575697824e-05
step: 320, loss: 4.0679202356841415e-05
step: 330, loss: 2.6473864636500366e-05
step: 340, loss: 2.1710315195377916e-05
step: 350, loss: 6.498570292023942e-05
step: 360, loss: 0.00012013353261863813
step: 370, loss: 0.00034540562774054706
step: 380, loss: 1.4278892194852233e-05
step: 390, loss: 8.779090421739966e-05
step: 400, loss: 1.3757345186604653e-05
step: 410, loss: 0.00015165629156399518
step: 420, loss: 2.0622759620891884e-05
step: 430, loss: 3.416682011447847e-05
step: 440, loss: 0.0005740467458963394
step: 450, loss: 0.0007064433302730322
step: 460, loss: 1.2155512195022311e-05
step: 470, loss: 0.0003259916848037392
step: 480, loss: 1.4111253221926745e-05
step: 490, loss: 0.001862179022282362
step: 500, loss: 3.0179611712810583e-05
step: 510, loss: 5.820859587402083e-05
step: 520, loss: 2.0034041881444864e-05
step: 530, loss: 2.4913613742683083e-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=0.9483960948396094, f1=0.9457579972183587,
best_f1=0.9447795823665893
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, 269.88it/s]
load_f1 = 0.9510945505356311
real_f1 = 0.9492314857941314
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, 268.38it/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 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: 0.8478099703788757

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

step: 10, loss: 0.06801556050777435

step: 20, loss: 0.38187217712402344

step: 30, loss: 0.36998337507247925

step: 40, loss: 0.4788874387741089

step: 50, loss: 0.27263376116752625

step: 60, loss: 0.3488416075706482

step: 70, loss: 0.18794621527194977

step: 80, loss: 0.3140389025211334

step: 90, loss: 0.3388538658618927

step: 100, loss: 0.09335856139659882

step: 110, loss: 0.3305206894874573

step: 120, loss: 0.18294519186019897

step: 130, loss: 0.1718929409980774

step: 140, loss: 0.19230325520038605

step: 150, loss: 0.1960202008485794

step: 160, loss: 0.18425965309143066

step: 170, loss: 0.062224797904491425

step: 180, loss: 0.17922250926494598

step: 190, loss: 0.2276221364736557

step: 200, loss: 0.17218798398971558

step: 210, loss: 0.35675084590911865

/usr/local/lib/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=0.6371308016877637, f1=0.6475770925110133,

best_f1=0.6475770925110133

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

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

step: 0, loss: 0.1297835260629654

step: 10, loss: 0.04882374778389931

step: 20, loss: 0.1732291281223297

step: 30, loss: 0.09587986022233963

step: 40, loss: 0.08215183019638062

step: 50, loss: 0.21289195120334625

step: 60, loss: 0.07067672908306122

step: 70, loss: 0.12016315013170242

step: 80, loss: 0.09186797589063644


```

step: 90, loss: 0.05104371905326843
step: 100, loss: 0.02884838730096817
step: 110, loss: 0.0645684003829956
step: 120, loss: 0.29349491000175476
step: 130, loss: 0.14355233311653137
step: 140, loss: 0.08498966693878174
step: 150, loss: 0.07585084438323975
step: 160, loss: 0.21637174487113953
step: 170, loss: 0.24259817600250244
step: 180, loss: 0.1731562614440918
step: 190, loss: 0.07536913454532623
step: 200, loss: 0.1092790961265564
step: 210, loss: 0.07403425872325897
/usr/local/lib/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=0.7178502879078695, f1=0.712, best_f1=0.712
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09337363392114639
step: 10, loss: 0.14292997121810913
step: 20, loss: 0.16540199518203735
step: 30, loss: 0.06456392258405685
step: 40, loss: 0.2250319868326187
step: 50, loss: 0.09235218167304993
step: 60, loss: 0.035823751240968704
step: 70, loss: 0.05117422342300415
step: 80, loss: 0.027603765949606895
step: 90, loss: 0.07119424641132355
step: 100, loss: 0.02114548347890377
step: 110, loss: 0.08318702131509781
step: 120, loss: 0.13675327599048615
step: 130, loss: 0.1844792664051056
step: 140, loss: 0.20880986750125885
step: 150, loss: 0.17834170162677765
step: 160, loss: 0.054467301815748215
step: 170, loss: 0.100193090736866
step: 180, loss: 0.00907951220870018
step: 190, loss: 0.2047155350446701
step: 200, loss: 0.09377282857894897
step: 210, loss: 0.1731633096933365
/usr/local/lib/python3.7/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=0.7140115163147792, f1=0.7109375, best_f1=0.712
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03341333195567131
step: 10, loss: 0.008717258460819721
step: 20, loss: 0.06998702883720398
step: 30, loss: 0.01665417291224003
step: 40, loss: 0.07273752987384796
step: 50, loss: 0.054885219782590866
step: 60, loss: 0.0532369427382946
step: 70, loss: 0.09172289818525314
step: 80, loss: 0.1575595736503601
step: 90, loss: 0.11205443739891052
step: 100, loss: 0.0620901845395565
step: 110, loss: 0.10422609746456146
step: 120, loss: 0.00324592599645257
step: 130, loss: 0.09866870939731598
step: 140, loss: 0.12775112688541412
step: 150, loss: 0.09040585160255432
step: 160, loss: 0.024699486792087555
step: 170, loss: 0.06948317587375641
step: 180, loss: 0.1497890204191208
step: 190, loss: 0.07985520362854004
step: 200, loss: 0.04042438790202141
step: 210, loss: 0.032894913107156754
/usr/local/lib/python3.7/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=0.709551656920078, f1=0.7134502923976609, best_f1=0.712
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03794388473033905
step: 10, loss: 0.06838948279619217
step: 20, loss: 0.034503038972616196
step: 30, loss: 0.056448064744472504
step: 40, loss: 0.11805649846792221
step: 50, loss: 0.0916917473077774
step: 60, loss: 0.03852378576993942
step: 70, loss: 0.0737944096326828
step: 80, loss: 0.016122620552778244
step: 90, loss: 0.04849879443645477
step: 100, loss: 0.06278938800096512
step: 110, loss: 0.01719038002192974
step: 120, loss: 0.007956981658935547
step: 130, loss: 0.031181151047348976

```

```

step: 140, loss: 0.0313493087887764
step: 150, loss: 0.022467875853180885
step: 160, loss: 0.01750563643872738
step: 170, loss: 0.011951049789786339
step: 180, loss: 0.09841432422399521
step: 190, loss: 0.07003077119588852
step: 200, loss: 0.10026342421770096
step: 210, loss: 0.011055907234549522
/usr/local/lib/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=0.744466800804829, f1=0.7185628742514971,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06025683134794235
step: 10, loss: 0.015111241489648819
step: 20, loss: 0.004950007423758507
step: 30, loss: 0.04951035976409912
step: 40, loss: 0.009632079862058163
step: 50, loss: 0.05115099623799324
step: 60, loss: 0.12498243153095245
step: 70, loss: 0.027009865269064903
step: 80, loss: 0.07853018492460251
step: 90, loss: 0.07313219457864761
step: 100, loss: 0.04385972023010254
step: 110, loss: 0.16799694299697876
step: 120, loss: 0.010825506411492825
step: 130, loss: 0.08288270235061646
step: 140, loss: 0.04637855663895607
step: 150, loss: 0.014640050940215588
step: 160, loss: 0.03519149497151375
step: 170, loss: 0.05626724660396576
step: 180, loss: 0.001990395365282893
step: 190, loss: 0.15239515900611877
step: 200, loss: 0.007901486940681934
step: 210, loss: 0.009208778850734234
/usr/local/lib/python3.7/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=0.7244094488188977, f1=0.7341269841269841,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0322246253490448
step: 10, loss: 0.014624390751123428
step: 20, loss: 0.005124703980982304
step: 30, loss: 0.08061407506465912
step: 40, loss: 0.041725099086761475
step: 50, loss: 0.021078946068882942
step: 60, loss: 0.12692567706108093
step: 70, loss: 0.011768486350774765
step: 80, loss: 0.045667264610528946
step: 90, loss: 0.007760465610772371
step: 100, loss: 0.019007178023457527
step: 110, loss: 0.04523225873708725
step: 120, loss: 0.044830434024333954
step: 130, loss: 0.0013847288209944963
step: 140, loss: 0.011995723471045494
step: 150, loss: 0.07960301637649536
step: 160, loss: 0.005629729945212603
step: 170, loss: 0.025734219700098038
step: 180, loss: 0.0007832664996385574
step: 190, loss: 0.0022307115141302347
step: 200, loss: 0.0024515583645552397
step: 210, loss: 0.034427519887685776
/usr/local/lib/python3.7/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=0.7342799188640974, f1=0.7321063394683026,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011275970377027988
step: 10, loss: 0.007906223647296429
step: 20, loss: 0.010886714793741703
step: 30, loss: 0.026682918891310692
step: 40, loss: 0.054074473679065704
step: 50, loss: 0.030074751004576683
step: 60, loss: 0.18608404695987701
step: 70, loss: 0.0012551896506920457
step: 80, loss: 0.0029738275334239006
step: 90, loss: 0.014450950548052788
step: 100, loss: 0.002075004857033491
step: 110, loss: 0.0021299482323229313
step: 120, loss: 0.002013616729527712
step: 130, loss: 0.004049450159072876
step: 140, loss: 0.004397721961140633
step: 150, loss: 0.04345685616135597

```

```

step: 160, loss: 0.07944801449775696
step: 170, loss: 0.006207325495779514
step: 180, loss: 0.013687401078641415
step: 190, loss: 0.00985369086265564
step: 200, loss: 0.01761983335018158
step: 210, loss: 0.16217821836471558
/usr/local/lib/python3.7/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=0.7241379310344829, f1=0.7121212121212123,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025787530466914177
step: 10, loss: 0.009495050646364689
step: 20, loss: 0.07507336884737015
step: 30, loss: 0.23313945531845093
step: 40, loss: 0.037786148488521576
step: 50, loss: 0.04768073558807373
step: 60, loss: 0.07254289835691452
step: 70, loss: 0.011985949240624905
step: 80, loss: 0.0016744460444897413
step: 90, loss: 0.0652012825012207
step: 100, loss: 0.007374117150902748
step: 110, loss: 0.003783909138292074
step: 120, loss: 0.0028788037598133087
step: 130, loss: 0.004291004966944456
step: 140, loss: 0.015211247839033604
step: 150, loss: 0.005472040269523859
step: 160, loss: 0.0003395032254047692
step: 170, loss: 0.019527100026607513
step: 180, loss: 0.043318942189216614
step: 190, loss: 0.13149131834506989
step: 200, loss: 0.013695796020328999
step: 210, loss: 0.08969353139400482
/usr/local/lib/python3.7/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=0.7276264591439688, f1=0.7283018867924528,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04037502780556679
step: 10, loss: 0.001206486951559782
step: 20, loss: 0.042638540267944336
step: 30, loss: 0.001750076306052506

```

```

step: 40, loss: 0.03321047127246857
step: 50, loss: 0.00246246624737978
step: 60, loss: 0.0002002313412958756
step: 70, loss: 0.007373875007033348
step: 80, loss: 8.662978507345542e-05
step: 90, loss: 0.036272525787353516
step: 100, loss: 0.00021609233226627111
step: 110, loss: 0.00017001786909531802
step: 120, loss: 0.0003887237689923495
step: 130, loss: 0.0006089008529670537
step: 140, loss: 0.0001503232924733311
step: 150, loss: 0.13483887910842896
step: 160, loss: 0.004482749849557877
step: 170, loss: 0.005112555809319019
step: 180, loss: 0.0007827744702808559
step: 190, loss: 0.10598386824131012
step: 200, loss: 0.06666423380374908
step: 210, loss: 0.005223384592682123
/usr/local/lib/python3.7/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=0.7438330170777988, f1=0.7362428842504745,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01908356137573719
step: 10, loss: 0.04175133630633354
step: 20, loss: 0.0030338545329868793
step: 30, loss: 0.0762556716799736
step: 40, loss: 0.012901579961180687
step: 50, loss: 0.0005937954992987216
step: 60, loss: 0.039258260279893875
step: 70, loss: 0.0011945280712097883
step: 80, loss: 0.024233564734458923
step: 90, loss: 0.0006606718525290489
step: 100, loss: 0.00048364081885665655
step: 110, loss: 0.015706147998571396
step: 120, loss: 0.000252056575845927
step: 130, loss: 0.02707895264029503
step: 140, loss: 0.012782365083694458
step: 150, loss: 0.0756540521979332
step: 160, loss: 0.00507877953350544
step: 170, loss: 0.07846065610647202
step: 180, loss: 0.11836300045251846
step: 190, loss: 0.021138619631528854
step: 200, loss: 0.0002571523655205965
step: 210, loss: 0.0006608176045119762

```

```

/usr/local/lib/python3.7/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=0.744466800804829, f1=0.7254509018036072,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00032679253490641713
step: 10, loss: 0.0009335947106592357
step: 20, loss: 0.0006207243422977626
step: 30, loss: 0.0012798907700926065
step: 40, loss: 0.0016206096624955535
step: 50, loss: 0.0072928536683321
step: 60, loss: 0.00074055977165699
step: 70, loss: 0.0002008676528930664
step: 80, loss: 0.00026566910673864186
step: 90, loss: 0.000710950989741832
step: 100, loss: 0.0016437626909464598
step: 110, loss: 0.0008951675263233483
step: 120, loss: 0.0024281537625938654
step: 130, loss: 0.020425301045179367
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 140, loss: 0.19964054226875305
step: 150, loss: 0.0005391223239712417
step: 160, loss: 0.004134295508265495
step: 170, loss: 0.004107173532247543
step: 180, loss: 0.00046344767906703055
step: 190, loss: 0.025855468586087227
step: 200, loss: 0.012029040604829788
step: 210, loss: 0.0008587630582042038
/usr/local/lib/python3.7/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=0.732919254658385, f1=0.7272727272727272,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00029565716977231205
step: 10, loss: 0.0003817765391431749
step: 20, loss: 0.00041783510823734105
step: 30, loss: 0.03921885043382645
step: 40, loss: 0.00020129478070884943
step: 50, loss: 0.0004107187851332128
step: 60, loss: 0.0004945165128447115
step: 70, loss: 0.013566295616328716
step: 80, loss: 0.0073348479345440865

```

```

step: 90, loss: 0.012143325991928577
step: 100, loss: 0.09990784525871277
step: 110, loss: 0.0008061345433816314
step: 120, loss: 0.007007574196904898
step: 130, loss: 0.001347996643744409
step: 140, loss: 0.01756748929619789
step: 150, loss: 0.00020513066556304693
step: 160, loss: 0.021954985335469246
step: 170, loss: 0.0037541536148637533
step: 180, loss: 0.02438906952738762
step: 190, loss: 6.788565224269405e-05
step: 200, loss: 0.0003955569409299642
step: 210, loss: 0.0003790140908677131
/usr/local/lib/python3.7/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=0.7272727272727274, f1=0.7319148936170214,
best_f1=0.7185628742514971
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0035878638736903667
step: 10, loss: 0.00027131338720209897
step: 20, loss: 0.0001012524007819593
step: 30, loss: 0.0002280031330883503
step: 40, loss: 0.0005114281084388494
step: 50, loss: 0.0007623145938850939
step: 60, loss: 0.0031383854802697897
step: 70, loss: 0.000782903574872762
step: 80, loss: 0.10480964928865433
step: 90, loss: 0.0422121062874794
step: 100, loss: 0.00028062897035852075
step: 110, loss: 0.00020170492643956095
step: 120, loss: 0.0296496469527483
step: 130, loss: 0.00019466593221295625
step: 140, loss: 0.00012955424608662724
step: 150, loss: 0.010452434420585632
step: 160, loss: 0.0010814788984134793
step: 170, loss: 0.00012462990707717836
step: 180, loss: 0.01767398789525032
step: 190, loss: 0.005825377535074949
step: 200, loss: 8.670879469718784e-05
step: 210, loss: 0.0010094528552144766
/usr/local/lib/python3.7/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=0.7344398340248962, f1=0.7272727272727272,
best_f1=0.7185628742514971

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008015169762074947
step: 10, loss: 0.002261560410261154
step: 20, loss: 0.0001285920588998124
step: 30, loss: 0.00017295803991146386
step: 40, loss: 0.0006148734828457236
step: 50, loss: 7.583006663480774e-05
step: 60, loss: 0.005572408903390169
step: 70, loss: 0.000157042799401097
step: 80, loss: 0.00028254278004169464
step: 90, loss: 0.024610944092273712
step: 100, loss: 0.00014006393030285835
step: 110, loss: 5.302035788190551e-05
step: 120, loss: 0.008041835390031338
step: 130, loss: 0.0018499484285712242
step: 140, loss: 0.0005941764102317393
step: 150, loss: 0.00039050407940521836
step: 160, loss: 0.013057572767138481
step: 170, loss: 0.006182391662150621
step: 180, loss: 0.01934407651424408
step: 190, loss: 0.00030388523009605706
step: 200, loss: 0.04172917827963829
step: 210, loss: 0.01868792064487934
/usr/local/lib/python3.7/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=0.7351129363449693, f1=0.7242798353909465,
best_f1=0.7185628742514971
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, 346.66it/s]
load_f1 = 0.7416173570019724
real_f1 = 0.7420634920634921
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, 256.05it/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 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: 0.866034746170044

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

step: 10, loss: 0.17258954048156738

step: 20, loss: 0.15328356623649597

step: 30, loss: 0.5159712433815002

step: 40, loss: 0.2579066753387451

step: 50, loss: 0.31158503890037537

```

step: 60, loss: 0.3575635552406311
step: 70, loss: 0.17731007933616638
step: 80, loss: 0.5200552940368652
step: 90, loss: 0.22737525403499603
step: 100, loss: 0.22040274739265442
step: 110, loss: 0.23514100909233093
step: 120, loss: 0.42884305119514465
step: 130, loss: 0.3393349349498749
step: 140, loss: 0.3338778614997864
step: 150, loss: 0.25684165954589844
step: 160, loss: 0.21786236763000488
step: 170, loss: 0.38134145736694336
step: 180, loss: 0.2970763146877289
step: 190, loss: 0.10671082884073257
/usr/local/lib/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=0.47858942065491183, f1=0.4708860759493671,
best_f1=0.4708860759493671
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2646031975746155
step: 10, loss: 0.07411599159240723
step: 20, loss: 0.054500892758369446
step: 30, loss: 0.19714610278606415
step: 40, loss: 0.4890337884426117
step: 50, loss: 0.3194114565849304
step: 60, loss: 0.11624795198440552
step: 70, loss: 0.33838513493537903
step: 80, loss: 0.18027782440185547
step: 90, loss: 0.12001896649599075
step: 100, loss: 0.27309438586235046
step: 110, loss: 0.12117460370063782
step: 120, loss: 0.18016384541988373
step: 130, loss: 0.09944133460521698
step: 140, loss: 0.2635049819946289
step: 150, loss: 0.011702113784849644
step: 160, loss: 0.06421803683042526
step: 170, loss: 0.17986007034778595
step: 180, loss: 0.17690375447273254
step: 190, loss: 0.09182973951101303
/usr/local/lib/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=0.7122507122507123, f1=0.712166172106825,
best_f1=0.712166172106825
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09519551694393158
step: 10, loss: 0.10043182969093323
step: 20, loss: 0.03631076589226723
step: 30, loss: 0.026740964502096176
step: 40, loss: 0.05443330854177475
step: 50, loss: 0.041805580258369446
step: 60, loss: 0.0444367378950119
step: 70, loss: 0.2257092148065567
step: 80, loss: 0.17043820023536682
step: 90, loss: 0.04653799161314964
step: 100, loss: 0.06763384491205215
step: 110, loss: 0.16243912279605865
step: 120, loss: 0.11044323444366455
step: 130, loss: 0.07517202943563461
step: 140, loss: 0.03207172453403473
step: 150, loss: 0.19006244838237762
step: 160, loss: 0.18665462732315063
step: 170, loss: 0.050736505538225174
step: 180, loss: 0.024986324831843376
step: 190, loss: 0.14172455668449402
/usr/local/lib/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=0.736, f1=0.7683923705722071, best_f1=0.7683923705722071
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07290089130401611
step: 10, loss: 0.022716253995895386
step: 20, loss: 0.011519001796841621
step: 30, loss: 0.016332030296325684
step: 40, loss: 0.023455634713172913
step: 50, loss: 0.012774639762938023
step: 60, loss: 0.09204273670911789

```

```

step: 70, loss: 0.034410398453474045
step: 80, loss: 0.08639050275087357
step: 90, loss: 0.0070191933773458
step: 100, loss: 0.03129751607775688
step: 110, loss: 0.006528459023684263
step: 120, loss: 0.016373304650187492
step: 130, loss: 0.31516575813293457
step: 140, loss: 0.05004153773188591
step: 150, loss: 0.032261498272418976
step: 160, loss: 0.07851094007492065
step: 170, loss: 0.012552632950246334
step: 180, loss: 0.1784411072731018
step: 190, loss: 0.05481487512588501
/usr/local/lib/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=0.7676767676767676, f1=0.7700258397932817,
best_f1=0.7700258397932817
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08402816951274872
step: 10, loss: 0.010721012018620968
step: 20, loss: 0.023725368082523346
step: 30, loss: 0.008220680989325047
step: 40, loss: 0.04124832898378372
step: 50, loss: 0.009004822000861168
step: 60, loss: 0.006416209973394871
step: 70, loss: 0.03800583258271217
step: 80, loss: 0.015606648288667202
step: 90, loss: 0.05382451415061951
step: 100, loss: 0.03479812294244766
step: 110, loss: 0.0034185945987701416
step: 120, loss: 0.0005810837028548121
step: 130, loss: 0.008983954787254333
step: 140, loss: 0.006762391421943903
step: 150, loss: 0.0028857423458248377
step: 160, loss: 0.001704955124296248
step: 170, loss: 0.014207742176949978
step: 180, loss: 0.023041822016239166
step: 190, loss: 0.07326526939868927
/usr/local/lib/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=0.7745098039215685, f1=0.7931034482758621,
best_f1=0.7931034482758621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024718858301639557
step: 10, loss: 0.00021742448734585196
step: 20, loss: 0.0011848786380141973
step: 30, loss: 0.0027241797652095556
step: 40, loss: 0.02973489835858345
step: 50, loss: 0.013723139651119709
step: 60, loss: 0.00029640397406183183
step: 70, loss: 0.007387886755168438
step: 80, loss: 0.008288716897368431
step: 90, loss: 0.001496052835136652
step: 100, loss: 0.0008601060835644603
step: 110, loss: 0.01351996697485447
step: 120, loss: 0.015423787757754326
step: 130, loss: 0.002779752714559436
step: 140, loss: 0.014594774693250656
step: 150, loss: 0.026983674615621567
step: 160, loss: 0.024670053273439407
step: 170, loss: 0.025967668741941452
step: 180, loss: 0.010733643546700478
step: 190, loss: 0.059372514486312866
/usr/local/lib/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=0.796875, f1=0.7676240208877285, best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038462679367512465
step: 10, loss: 0.07451833784580231
step: 20, loss: 0.0003123946371488273
step: 30, loss: 0.006025644484907389
step: 40, loss: 0.012216538190841675
step: 50, loss: 0.22442762553691864
step: 60, loss: 0.006012361962348223
step: 70, loss: 0.0021765008568763733

```

```

step: 80, loss: 0.035479769110679626
step: 90, loss: 0.007576988078653812
step: 100, loss: 0.0023397912736982107
step: 110, loss: 0.020517533645033836
step: 120, loss: 0.016263844445347786
step: 130, loss: 0.00263204169459641
step: 140, loss: 0.0005980654968880117
step: 150, loss: 0.028002819046378136
step: 160, loss: 0.0004843634960707277
step: 170, loss: 0.0027708844281733036
step: 180, loss: 0.04359782114624977
step: 190, loss: 0.001329035614617169
/usr/local/lib/python3.7/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=0.7647058823529412, f1=0.784741144414169,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005649367230944335
step: 10, loss: 0.0014339471235871315
step: 20, loss: 0.0013267196482047439
step: 30, loss: 0.0449419803917408
step: 40, loss: 0.002903795801103115
step: 50, loss: 0.002096920507028699
step: 60, loss: 0.0007339322473853827
step: 70, loss: 0.08290833979845047
step: 80, loss: 0.022255852818489075
step: 90, loss: 0.0013166337739676237
step: 100, loss: 0.02328961342573166
step: 110, loss: 0.0018084567273035645
step: 120, loss: 0.005917919334024191
step: 130, loss: 0.002743019489571452
step: 140, loss: 0.0020954005885869265
step: 150, loss: 0.0845620185136795
step: 160, loss: 0.0007491537253372371
step: 170, loss: 0.000192764462553896
step: 180, loss: 0.004755682311952114
step: 190, loss: 0.0028554839082062244
/usr/local/lib/python3.7/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=0.7673860911270982, f1=0.7949367088607595,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.013566742651164532
step: 10, loss: 0.015231005847454071
step: 20, loss: 0.004702767822891474
step: 30, loss: 0.0012284325202926993
step: 40, loss: 0.0011655634734779596
step: 50, loss: 0.0004066956462338567
step: 60, loss: 0.00046057457802817225
step: 70, loss: 0.0018778087105602026
step: 80, loss: 0.0003460564184933901
step: 90, loss: 0.00497980322688818
step: 100, loss: 0.0005522360443137586
step: 110, loss: 0.04330601543188095
step: 120, loss: 0.004478972405195236
step: 130, loss: 0.0072077056393027306
step: 140, loss: 0.00032169235055334866
step: 150, loss: 0.13340353965759277
step: 160, loss: 0.0025634728372097015
step: 170, loss: 0.0007508920971304178
step: 180, loss: 0.0052962526679039
step: 190, loss: 0.0002624971675686538
/usr/local/lib/python3.7/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=0.7724867724867726, f1=0.8116710875331565,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033334983163513243
step: 10, loss: 0.0011491953628137708
step: 20, loss: 0.0003027615894097835
step: 30, loss: 0.00041339005110785365
step: 40, loss: 0.00027513582608662546
step: 50, loss: 0.0008937659440562129
step: 60, loss: 0.0032021277584135532
step: 70, loss: 0.0016528369160369039
step: 80, loss: 0.0001455436577089131
step: 90, loss: 0.0006654388271272182
step: 100, loss: 0.0007490653661079705
step: 110, loss: 0.0004402115009725094
step: 120, loss: 0.004613331984728575
step: 130, loss: 0.0028319957200437784
step: 140, loss: 0.000396536139305681
step: 150, loss: 0.00018186570378020406
step: 160, loss: 0.0001982853136723861
step: 170, loss: 0.0009255098993889987
step: 180, loss: 0.0005010427557863295
step: 190, loss: 0.0014839783543720841

```

```

/usr/local/lib/python3.7/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=0.7810945273631841, f1=0.785, best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017152976943179965
step: 10, loss: 0.0013995092594996095
step: 20, loss: 0.002945970045402646
step: 30, loss: 0.01654825359582901
step: 40, loss: 0.0004452514695003629
step: 50, loss: 0.00022663090203423053
step: 60, loss: 0.0005127665353938937
step: 70, loss: 0.00022449324023909867
step: 80, loss: 0.0004493400047067553
step: 90, loss: 0.0006123529747128487
step: 100, loss: 0.00015138232265599072
step: 110, loss: 9.408251207787544e-05
step: 120, loss: 0.00042790407314896584
step: 130, loss: 0.0015975175192579627
step: 140, loss: 0.0004017602186650038
step: 150, loss: 0.00081476173363626
step: 160, loss: 0.0001950799342012033
step: 170, loss: 0.00023147034517023712
step: 180, loss: 0.001190291834063828
step: 190, loss: 0.00033616588916629553
/usr/local/lib/python3.7/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=0.7780548628428927, f1=0.7905759162303665,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00076251698192209
step: 10, loss: 0.05681851878762245
step: 20, loss: 0.00044444508966989815
step: 30, loss: 0.0008593123639002442
step: 40, loss: 0.00031391900847665966
step: 50, loss: 0.002548250835388899
step: 60, loss: 0.0001272221707040444
step: 70, loss: 0.00019119489297736436
step: 80, loss: 0.0001719934807624668
step: 90, loss: 0.004851853474974632
step: 100, loss: 0.0026062375400215387
step: 110, loss: 0.0012307371944189072
step: 120, loss: 0.0001297235139645636

```

```

step: 130, loss: 0.0001610197505215183
step: 140, loss: 0.0008609868818894029
step: 150, loss: 0.00027172997943125665
step: 160, loss: 0.00038674776442348957
step: 170, loss: 0.0014197032433003187
step: 180, loss: 0.00013235464575700462
step: 190, loss: 0.033371537923812866
/usr/local/lib/python3.7/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=0.7897435897435896, f1=0.8167539267015705,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009698187932372093
step: 10, loss: 0.0004763129400089383
step: 20, loss: 0.15317347645759583
step: 30, loss: 0.0002214101405115798
step: 40, loss: 0.000151043786900118
step: 50, loss: 0.0002240701433038339
step: 60, loss: 0.00023934096680022776
step: 70, loss: 0.00017545692389830947
step: 80, loss: 0.0004666441527660936
step: 90, loss: 0.0015666306717321277
step: 100, loss: 0.0005535543314181268
step: 110, loss: 0.0003183176158927381
step: 120, loss: 0.012624109163880348
step: 130, loss: 0.00039173694676719606
step: 140, loss: 0.005491478368639946
step: 150, loss: 0.0002626345958560705
step: 160, loss: 0.0009725496056489646
step: 170, loss: 0.00025292401551268995
step: 180, loss: 0.0015726598212495446
step: 190, loss: 0.0003005262406077236
/usr/local/lib/python3.7/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=0.7860696517412935, f1=0.8062015503875969,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01826508529484272
step: 10, loss: 0.0003799505066126585
step: 20, loss: 0.00023756088921800256
step: 30, loss: 0.02154596894979477
step: 40, loss: 0.0002036722144111991

```

```

step: 50, loss: 0.002353039337322116
step: 60, loss: 9.360387048218399e-05
step: 70, loss: 0.00046203500824049115
step: 80, loss: 6.907425995450467e-05
step: 90, loss: 0.00016283054719679058
step: 100, loss: 0.00017855459009297192
step: 110, loss: 0.0018902318552136421
step: 120, loss: 0.000740925723221153
step: 130, loss: 6.054401455912739e-05
step: 140, loss: 0.00016025487275328487
step: 150, loss: 0.00013860390754416585
step: 160, loss: 0.031093111261725426
step: 170, loss: 0.00033039890695363283
step: 180, loss: 7.762081077089533e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.0001382004120387137
/usr/local/lib/python3.7/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=0.7814207650273224, f1=0.7774647887323942,
best_f1=0.7676240208877285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001133655387093313
step: 10, loss: 0.00018557642761152238
step: 20, loss: 0.0005122090224176645
step: 30, loss: 0.001235966570675373
step: 40, loss: 0.0006135939038358629
step: 50, loss: 0.0001526160485809669
step: 60, loss: 0.005018001422286034
step: 70, loss: 0.0004670402850024402
step: 80, loss: 0.0026599676348268986
step: 90, loss: 0.0007646363810636103
step: 100, loss: 0.00015192307182587683
step: 110, loss: 0.00018342750263400376
step: 120, loss: 0.00010680089326342568
step: 130, loss: 0.0010087505215778947
step: 140, loss: 0.0006356005906127393
step: 150, loss: 0.0025553149171173573
step: 160, loss: 0.0011378376511856914
step: 170, loss: 0.0002998167765326798
step: 180, loss: 0.026823600754141808
step: 190, loss: 0.011337579227983952
/usr/local/lib/python3.7/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=0.7807486631016043, f1=0.7835616438356164,

```

```
best_f1=0.7676240208877285
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.")
```

```
2049it [00:08, 231.00it/s]
```

```
load_f1 = 0.7882037533512065
```

```
real_f1 = 0.7862796833773087
```

```
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, 263.97it/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: 0.8464320302009583
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22633279860019684
step: 20, loss: 0.153848797082901
step: 30, loss: 0.23070599138736725
step: 40, loss: 0.31368720531463623
step: 50, loss: 0.37344348430633545
step: 60, loss: 0.43824225664138794
step: 70, loss: 0.2985379099845886
step: 80, loss: 0.2561792731285095
step: 90, loss: 0.3766288459300995
step: 100, loss: 0.220526784658432
step: 110, loss: 0.18663954734802246
step: 120, loss: 0.5624297261238098
step: 130, loss: 0.3990207612514496
step: 140, loss: 0.4876222610473633
step: 150, loss: 0.12739169597625732
step: 160, loss: 0.3295294940471649
step: 170, loss: 0.1917489618062973
/usr/local/lib/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=0.5587467362924283, f1=0.5454545454545455,
best_f1=0.5454545454545455
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.36517342925071716
step: 10, loss: 0.0905388742685318
step: 20, loss: 0.33609122037887573

```

```

step: 30, loss: 0.23064154386520386
step: 40, loss: 0.09247549623250961
step: 50, loss: 0.16869129240512848
step: 60, loss: 0.06647677719593048
step: 70, loss: 0.21230517327785492
step: 80, loss: 0.13936446607112885
step: 90, loss: 0.1520930379629135
step: 100, loss: 0.10163440555334091
step: 110, loss: 0.19240282475948334
step: 120, loss: 0.038976021111011505
step: 130, loss: 0.06012362986803055
step: 140, loss: 0.09713274240493774
step: 150, loss: 0.0721026211977005
step: 160, loss: 0.022406762465834618
step: 170, loss: 0.01658041588962078
/usr/local/lib/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=0.7864077669902912, f1=0.7901785714285714,
best_f1=0.7901785714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03733387961983681
step: 10, loss: 0.07187020033597946
step: 20, loss: 0.006441513542085886
step: 30, loss: 0.3707144856452942
step: 40, loss: 0.04783328250050545
step: 50, loss: 0.06486669927835464
step: 60, loss: 0.06855856627225876
step: 70, loss: 0.06297533214092255
step: 80, loss: 0.13067741692066193
step: 90, loss: 0.05797741562128067
step: 100, loss: 0.03202669695019722
step: 110, loss: 0.13867689669132233
step: 120, loss: 0.0030584512278437614
step: 130, loss: 0.08532708883285522
step: 140, loss: 0.015575256198644638
step: 150, loss: 0.02053629234433174
step: 160, loss: 0.004894785117357969
step: 170, loss: 0.16180363297462463
/usr/local/lib/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=0.8146453089244851, f1=0.7922912205567453,
best_f1=0.7922912205567453
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06584277004003525
step: 10, loss: 0.0681089386343956
step: 20, loss: 0.00995376706123352
step: 30, loss: 0.014700713567435741
step: 40, loss: 0.006602626759558916
step: 50, loss: 0.011534737423062325
step: 60, loss: 0.016164489090442657
step: 70, loss: 0.013055272400379181
step: 80, loss: 0.009649310261011124
step: 90, loss: 0.01664401777088642
step: 100, loss: 0.006880162749439478
step: 110, loss: 0.08399064093828201
step: 120, loss: 0.14707346260547638
step: 130, loss: 0.04542073607444763
step: 140, loss: 0.0031604073010385036
step: 150, loss: 0.0019985847175121307
step: 160, loss: 0.020849479362368584
step: 170, loss: 0.01596442423760891
/usr/local/lib/python3.7/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=0.810304449648712, f1=0.8054298642533937,
best_f1=0.7922912205567453
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013923263177275658
step: 10, loss: 0.013363362289965153
step: 20, loss: 0.0059830620884895325
step: 30, loss: 0.008932424709200859
step: 40, loss: 0.00555869797244668
step: 50, loss: 0.012666813097894192
step: 60, loss: 0.009628557600080967
step: 70, loss: 0.030987929552793503
step: 80, loss: 0.003788159228861332
step: 90, loss: 0.0064508942887187
step: 100, loss: 0.012716082856059074
step: 110, loss: 0.04778299480676651
step: 120, loss: 0.04149126634001732

```

```

step: 130, loss: 0.0005932598141953349
step: 140, loss: 0.02230343408882618
step: 150, loss: 0.014727208763360977
step: 160, loss: 0.005737625993788242
step: 170, loss: 0.015876246616244316
/usr/local/lib/python3.7/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=0.803532008830022, f1=0.778494623655914,
best_f1=0.7922912205567453
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012175717391073704
step: 10, loss: 0.047821562737226486
step: 20, loss: 0.002069673500955105
step: 30, loss: 0.008926721289753914
step: 40, loss: 0.0013549122959375381
step: 50, loss: 0.003126419149339199
step: 60, loss: 0.07235712558031082
step: 70, loss: 0.018444301560521126
step: 80, loss: 0.016688862815499306
step: 90, loss: 0.032432086765766144
step: 100, loss: 0.03679336607456207
step: 110, loss: 0.001581057207658887
step: 120, loss: 0.05807045102119446
step: 130, loss: 0.19377346336841583
step: 140, loss: 0.0014753271825611591
step: 150, loss: 0.10180415213108063
step: 160, loss: 0.09582267701625824
step: 170, loss: 0.00258841086179018
/usr/local/lib/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=0.815165876777251, f1=0.7927927927927928,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001600712537765503
step: 10, loss: 0.0005521973944269121
step: 20, loss: 0.0006791295600123703
step: 30, loss: 0.0009274943149648607
step: 40, loss: 0.011491443030536175

```

```

step: 50, loss: 0.0006297604995779693
step: 60, loss: 0.11823201179504395
step: 70, loss: 0.0014776478055864573
step: 80, loss: 0.0020559753756970167
step: 90, loss: 0.009543842636048794
step: 100, loss: 0.02449030801653862
step: 110, loss: 0.0035847043618559837
step: 120, loss: 0.0649237409234047
step: 130, loss: 0.03374312072992325
step: 140, loss: 0.020047489553689957
step: 150, loss: 0.014971405267715454
step: 160, loss: 0.0165573563426733
step: 170, loss: 0.147870272397995
/usr/local/lib/python3.7/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=0.7871853546910755, f1=0.7695652173913043,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0374046191573143
step: 10, loss: 0.06438684463500977
step: 20, loss: 0.0058878641575574875
step: 30, loss: 0.003822281025350094
step: 40, loss: 0.0007854446303099394
step: 50, loss: 0.0005649691447615623
step: 60, loss: 0.006923405919224024
step: 70, loss: 0.0280456505715847
step: 80, loss: 0.002132520778104663
step: 90, loss: 0.001136253820732236
step: 100, loss: 0.010903437621891499
step: 110, loss: 0.03867195546627045
step: 120, loss: 0.0009668486891314387
step: 130, loss: 0.0012081189779564738
step: 140, loss: 0.0005772241274826229
step: 150, loss: 0.03598568215966225
step: 160, loss: 0.06423579156398773
step: 170, loss: 0.003154386067762971
/usr/local/lib/python3.7/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=0.8061002178649237, f1=0.7914893617021277,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000979239703156054

```

```

step: 10, loss: 0.0019243636634200811
step: 20, loss: 0.00029805843951180577
step: 30, loss: 0.0018902316223829985
step: 40, loss: 0.008377520367503166
step: 50, loss: 0.0047103785909712315
step: 60, loss: 0.0007137761567719281
step: 70, loss: 0.06106722727417946
step: 80, loss: 0.06572006642818451
step: 90, loss: 0.014741514809429646
step: 100, loss: 0.03784267231822014
step: 110, loss: 0.0002746776444837451
step: 120, loss: 0.0030782329849898815
step: 130, loss: 0.004400031175464392
step: 140, loss: 0.0014081436675041914
step: 150, loss: 0.012444366700947285
step: 160, loss: 0.012216150760650635
step: 170, loss: 0.008067392744123936
/usr/local/lib/python3.7/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=0.8020304568527918, f1=0.8240963855421687,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008288141689263284
step: 10, loss: 0.005485715344548225
step: 20, loss: 0.00020190364739391953
step: 30, loss: 0.00016357458662241697
step: 40, loss: 0.001295819878578186
step: 50, loss: 0.0016454914584755898
step: 60, loss: 0.00030680643976666033
step: 70, loss: 0.0002741374773904681
step: 80, loss: 0.10995826870203018
step: 90, loss: 0.016977041959762573
step: 100, loss: 0.021737925708293915
step: 110, loss: 0.002519806381314993
step: 120, loss: 0.013948807492852211
step: 130, loss: 0.005586464423686266
step: 140, loss: 0.0021778051741421223
step: 150, loss: 0.0006908076466061175
step: 160, loss: 8.703768253326416e-05
step: 170, loss: 0.006890503689646721
/usr/local/lib/python3.7/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=0.8051282051282052, f1=0.8321167883211679,
best_f1=0.7927927927927928

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011439025547588244
step: 10, loss: 0.03640966862440109
step: 20, loss: 0.05451758950948715
step: 30, loss: 0.0002798606001306325
step: 40, loss: 0.0027118725702166557
step: 50, loss: 0.005313791334629059
step: 60, loss: 0.00029538394301198423
step: 70, loss: 0.0001017041431623511
step: 80, loss: 0.00017888662114273757
step: 90, loss: 0.001503828214481473
step: 100, loss: 0.0002505419834051281
step: 110, loss: 0.001904781791381538
step: 120, loss: 0.0030437526293098927
step: 130, loss: 0.000275795697234571
step: 140, loss: 0.01143950317054987
step: 150, loss: 0.0002499337715562433
step: 160, loss: 0.006819857284426689
step: 170, loss: 0.051945898681879044
/usr/local/lib/python3.7/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=0.7851851851851852, f1=0.8074245939675173,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0056240614503622055
step: 10, loss: 0.0020444137044250965
step: 20, loss: 0.014969246461987495
step: 30, loss: 0.0001341060851700604
step: 40, loss: 0.000117207266157493
step: 50, loss: 0.0001518816570751369
step: 60, loss: 0.00018618410103954375
step: 70, loss: 0.010428049601614475
step: 80, loss: 0.08943653851747513
step: 90, loss: 9.409039921592921e-05
step: 100, loss: 8.788728882791474e-05
step: 110, loss: 0.00011038436059607193
step: 120, loss: 0.00010394455603091046
step: 130, loss: 0.00012799282558262348
step: 140, loss: 0.00015354796778410673
step: 150, loss: 0.0008467122679576278
step: 160, loss: 0.0006261161179281771
step: 170, loss: 0.002758581656962633
/usr/local/lib/python3.7/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=0.7874015748031495, f1=0.8195121951219512,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017396461043972522
step: 10, loss: 0.040945567190647125
step: 20, loss: 0.0007361001335084438
step: 30, loss: 8.960995182860643e-05
step: 40, loss: 6.434662645915523e-05
step: 50, loss: 0.004812205210328102
step: 60, loss: 0.00017841870430856943
step: 70, loss: 0.019959360361099243
step: 80, loss: 0.0014174024108797312
step: 90, loss: 0.00014510068285744637
step: 100, loss: 7.75712396716699e-05
step: 110, loss: 0.0002747728431131691
step: 120, loss: 0.0002558643463999033
step: 130, loss: 0.0029480690136551857
step: 140, loss: 7.386572542600334e-05
step: 150, loss: 0.0007762728491798043
step: 160, loss: 9.815625526243821e-05
step: 170, loss: 0.00010122166713699698
/usr/local/lib/python3.7/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=0.7905759162303666, f1=0.8235294117647058,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011817395716207102
step: 10, loss: 6.682154344161972e-05
step: 20, loss: 0.0004720633151009679
step: 30, loss: 0.009890713728964329
step: 40, loss: 0.00019755212997552007
step: 50, loss: 0.000647631473839283
step: 60, loss: 0.00011105719750048593
step: 70, loss: 0.000225688680075109
step: 80, loss: 0.005172157660126686
step: 90, loss: 0.00010977822967106476
step: 100, loss: 0.000708036997821182
step: 110, loss: 0.00034925361978821456
step: 120, loss: 0.0006850342615507543
step: 130, loss: 0.00019607938884291798
step: 140, loss: 0.0005360357463359833

```

```

step: 150, loss: 0.00016731290088500828
step: 160, loss: 0.034059859812259674
step: 170, loss: 0.008218168281018734
/usr/local/lib/python3.7/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=0.7939698492462312, f1=0.8132387706855791,
best_f1=0.7927927927927928
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014936257502995431
step: 10, loss: 0.0068393247202038765
step: 20, loss: 0.004203058313578367
step: 30, loss: 0.055208269506692886
step: 40, loss: 8.548861660528928e-05
step: 50, loss: 7.25904101273045e-05
step: 60, loss: 6.720957753714174e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.13273733854293823
step: 80, loss: 8.849246660247445e-05
step: 90, loss: 9.33401970542036e-05
step: 100, loss: 0.0003948849334847182
step: 110, loss: 0.00015015690587460995
step: 120, loss: 0.03655868023633957
step: 130, loss: 0.0026086659636348486
step: 140, loss: 8.999095734907314e-05
step: 150, loss: 0.07163415849208832
step: 160, loss: 8.809621067484841e-05
step: 170, loss: 0.001949519501067698
/usr/local/lib/python3.7/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=0.7929292929292929, f1=0.8104265402843601,
best_f1=0.7927927927927928
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, 410.94it/s]

load_f1 = 0.8142857142857144

real_f1 = 0.8104265402843601

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 [00:12, 347.59it/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 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: 0.8167072534561157
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.46241694688796997
step: 20, loss: 0.5833951830863953
step: 30, loss: 0.4091501832008362
step: 40, loss: 0.16794539988040924
step: 50, loss: 0.09512589871883392
step: 60, loss: 0.07601802796125412
step: 70, loss: 0.11679311841726303
step: 80, loss: 0.12787838280200958
step: 90, loss: 0.009269451722502708
step: 100, loss: 0.11162637174129486
step: 110, loss: 0.04144415259361267
step: 120, loss: 0.005820597987622023
step: 130, loss: 0.005109719466418028
step: 140, loss: 0.005522135179489851
step: 150, loss: 0.14811928570270538
step: 160, loss: 0.1573498696088791
step: 170, loss: 0.016914859414100647
step: 180, loss: 0.011391712352633476
step: 190, loss: 0.06058811396360397
step: 200, loss: 0.008924082852900028
step: 210, loss: 0.0046596345491707325
step: 220, loss: 0.0030827417504042387
step: 230, loss: 0.004870019853115082
/usr/local/lib/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=0.9566184649610678,
best_f1=0.9566184649610678
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007075534202158451
step: 10, loss: 0.0011350710410624743
step: 20, loss: 0.01303175836801529
step: 30, loss: 0.002615845063701272
step: 40, loss: 0.018783599138259888
step: 50, loss: 0.046720534563064575
step: 60, loss: 0.007782787550240755
step: 70, loss: 0.023637229576706886
step: 80, loss: 0.0005233291303738952
step: 90, loss: 0.0031308161560446024
step: 100, loss: 0.12330637127161026

```

```

step: 110, loss: 0.11770768463611603
step: 120, loss: 0.0019127195701003075
step: 130, loss: 0.007674718741327524
step: 140, loss: 0.34375885128974915
step: 150, loss: 0.03098350018262863
step: 160, loss: 0.005000862758606672
step: 170, loss: 0.011823413893580437
step: 180, loss: 0.002984385471791029
step: 190, loss: 0.07805491238832474
step: 200, loss: 0.006470758002251387
step: 210, loss: 0.06421292573213577
step: 220, loss: 0.004207639489322901
step: 230, loss: 0.009151873178780079
/usr/local/lib/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=0.9787234042553192, f1=0.9696287964004499,
best_f1=0.9696287964004499
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10065848380327225
step: 10, loss: 0.0059197130613029
step: 20, loss: 0.011979926377534866
step: 30, loss: 0.036534786224365234
step: 40, loss: 0.006625022739171982
step: 50, loss: 0.00374836684204638
step: 60, loss: 0.0007576234056614339
step: 70, loss: 0.0030714652966707945
step: 80, loss: 0.005092759616672993
step: 90, loss: 0.006524829193949699
step: 100, loss: 0.0005217724828980863
step: 110, loss: 0.0008819756330922246
step: 120, loss: 0.0011512050405144691
step: 130, loss: 0.0005305457743816078
step: 140, loss: 0.2528841495513916
step: 150, loss: 0.000736319285351783
step: 160, loss: 0.003574762726202607
step: 170, loss: 0.005906072445213795
step: 180, loss: 0.003326411359012127
step: 190, loss: 0.0011868157889693975
step: 200, loss: 0.009281323291361332
step: 210, loss: 0.03731123358011246
step: 220, loss: 0.008505190722644329

```

```

step: 230, loss: 0.01779237948358059
/usr/local/lib/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=0.9788182831661093, f1=0.9731543624161074,
best_f1=0.9731543624161074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004863814916461706
step: 10, loss: 0.001560900709591806
step: 20, loss: 0.0003505370987113565
step: 30, loss: 0.0014455848140642047
step: 40, loss: 0.012001906521618366
step: 50, loss: 0.0017290284158661962
step: 60, loss: 0.0006083257612772286
step: 70, loss: 0.010256837122142315
step: 80, loss: 0.06632440537214279
step: 90, loss: 0.003920251037925482
step: 100, loss: 0.00372475222684443
step: 110, loss: 0.013339745812118053
step: 120, loss: 0.012505949474871159
step: 130, loss: 0.009192192927002907
step: 140, loss: 0.00023315528233069927
step: 150, loss: 0.0006528829107992351
step: 160, loss: 0.0005917042144574225
step: 170, loss: 0.015474751591682434
step: 180, loss: 0.08982416987419128
step: 190, loss: 0.007313722278922796
step: 200, loss: 0.0013497992185875773
step: 210, loss: 0.009198810905218124
step: 220, loss: 0.0007960486691445112
step: 230, loss: 0.0004554597835522145
/usr/local/lib/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=0.9810901001112348, f1=0.974472807991121,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007664929144084454
step: 10, loss: 0.0017688384978100657
step: 20, loss: 0.0004847165837418288
step: 30, loss: 0.00025068887043744326
step: 40, loss: 0.0003056653658859432
step: 50, loss: 0.00037387176416814327
step: 60, loss: 0.0022630849853157997
step: 70, loss: 0.0005633901455439627
step: 80, loss: 0.0011824864195659757
step: 90, loss: 0.003465858055278659
step: 100, loss: 0.002772532170638442
step: 110, loss: 0.002665579319000244
step: 120, loss: 0.051970209926366806
step: 130, loss: 0.03552548959851265
step: 140, loss: 0.0006182768847793341
step: 150, loss: 0.00014368562551680952
step: 160, loss: 0.026463687419891357
step: 170, loss: 0.027844997122883797
step: 180, loss: 0.0005161516601219773
step: 190, loss: 0.0009612050489522517
step: 200, loss: 0.0005135826067999005
step: 210, loss: 0.0003236384363844991
step: 220, loss: 0.00046724145067855716
step: 230, loss: 0.001389609300531447
/usr/local/lib/python3.7/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=0.9789590254706534, f1=0.9679558011049725,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017600316787138581
step: 10, loss: 0.00026205528411082923
step: 20, loss: 0.001211695373058319
step: 30, loss: 0.0017181998118758202
step: 40, loss: 0.0023553245700895786
step: 50, loss: 0.0797281488776207
step: 60, loss: 0.009659284725785255
step: 70, loss: 0.0010356752900406718
step: 80, loss: 0.0002938808174803853
step: 90, loss: 0.0008326725219376385
step: 100, loss: 0.00026404226082377136
step: 110, loss: 0.03326454386115074
step: 120, loss: 0.00026028527645394206
step: 130, loss: 0.005840679164975882
step: 140, loss: 0.0025617340579628944

```

```

step: 150, loss: 0.00034036656143143773
step: 160, loss: 0.03616759181022644
step: 170, loss: 0.0025206191930919886
step: 180, loss: 0.0008908042800612748
step: 190, loss: 0.0003871851949952543
step: 200, loss: 0.008679709397256374
step: 210, loss: 0.0009209315176121891
step: 220, loss: 0.0002627809881232679
step: 230, loss: 0.0003432459197938442
/usr/local/lib/python3.7/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=0.9755011135857461, f1=0.9745293466223698,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15374770760536194
step: 10, loss: 0.0001305479818256572
step: 20, loss: 0.00027009673067368567
step: 30, loss: 0.0029432072769850492
step: 40, loss: 0.002887460170313716
step: 50, loss: 0.00011875689233420417
step: 60, loss: 0.010345268063247204
step: 70, loss: 0.00031984414090402424
step: 80, loss: 0.0001944906689459458
step: 90, loss: 0.00032605757587589324
step: 100, loss: 0.0002875561476685107
step: 110, loss: 0.0010131074814125896
step: 120, loss: 0.0006505827768705785
step: 130, loss: 0.009111128747463226
step: 140, loss: 9.219637286150828e-05
step: 150, loss: 0.0001540674566058442
step: 160, loss: 0.00018472829833626747
step: 170, loss: 0.00015181991329882294
step: 180, loss: 0.0023583443835377693
step: 190, loss: 0.000583296234253794
step: 200, loss: 0.00042514901724644005
step: 210, loss: 0.00014304592332337052
step: 220, loss: 0.0017934143543243408
step: 230, loss: 0.006033156532794237
/usr/local/lib/python3.7/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=0.9799107142857142, f1=0.9700996677740864,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013361285673454404
step: 10, loss: 0.10677609592676163
step: 20, loss: 0.0002202207106165588
step: 30, loss: 0.00017346808454021811
step: 40, loss: 0.00016030119149945676
step: 50, loss: 0.010680940933525562
step: 60, loss: 0.00378697132691741
step: 70, loss: 0.00014023066614754498
step: 80, loss: 0.000897403631825
step: 90, loss: 0.000235116618569009
step: 100, loss: 0.0008548255427740514
step: 110, loss: 0.00014284433564171195
step: 120, loss: 9.695907647255808e-05
step: 130, loss: 0.0957895815372467
step: 140, loss: 0.00010502694203751162
step: 150, loss: 0.0007954547181725502
step: 160, loss: 0.003613205160945654
step: 170, loss: 0.0010044642258435488
step: 180, loss: 0.004451517015695572
step: 190, loss: 0.001993349054828286
step: 200, loss: 0.0034044256899505854
step: 210, loss: 0.0001263837330043316
step: 220, loss: 0.0006240012007765472
step: 230, loss: 0.041796911507844925
/usr/local/lib/python3.7/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=0.9753363228699552, f1=0.9743016759776536,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008065198548138142
step: 10, loss: 0.00045150815276429057
step: 20, loss: 0.0008959250990301371
step: 30, loss: 0.007069804705679417
step: 40, loss: 9.060029697138816e-05
step: 50, loss: 0.00025342905428260565
step: 60, loss: 0.0002671517140697688
step: 70, loss: 0.00015557228471152484
step: 80, loss: 0.05427694693207741
step: 90, loss: 0.0007553371833637357
step: 100, loss: 0.003411377314478159
step: 110, loss: 0.0005014523630961776
step: 120, loss: 0.06776349991559982
step: 130, loss: 0.00010720917634898797
step: 140, loss: 0.00019909386173821986

```

```

step: 150, loss: 5.8184523368254304e-05
step: 160, loss: 0.0001293062960030511
step: 170, loss: 5.7832956372294575e-05
step: 180, loss: 0.0006711662863381207
step: 190, loss: 0.00010065750393550843
step: 200, loss: 0.0024690604768693447
step: 210, loss: 0.00021203081996645778
step: 220, loss: 0.02634068951010704
step: 230, loss: 0.005826334934681654
/usr/local/lib/python3.7/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=0.9720670391061451, f1=0.9755011135857461,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12067167460918427
step: 10, loss: 5.902999691898003e-05
step: 20, loss: 0.00011199377331649885
step: 30, loss: 0.0005090002086944878
step: 40, loss: 0.00013581343227997422
step: 50, loss: 0.0006500667659565806
step: 60, loss: 0.03933987766504288
step: 70, loss: 0.00019896433514077216
step: 80, loss: 0.00044816179433837533
step: 90, loss: 0.00019373782561160624
step: 100, loss: 0.0005011777975596488
step: 110, loss: 0.0002531710488256067
step: 120, loss: 0.017078237608075142
step: 130, loss: 0.0013504693051800132
step: 140, loss: 0.020472409203648567
step: 150, loss: 7.700661808485165e-05
step: 160, loss: 0.00010677809768822044
step: 170, loss: 9.16684148251079e-05
step: 180, loss: 0.0001545526902191341
step: 190, loss: 0.0008194343536160886
step: 200, loss: 7.868100510677323e-05
step: 210, loss: 7.17955845175311e-05
step: 220, loss: 0.001596952206455171
step: 230, loss: 0.0006492432439699769
/usr/local/lib/python3.7/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=0.9776785714285714, f1=0.9766925638179801,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.482363434974104e-05
step: 10, loss: 0.003665438387542963
step: 20, loss: 3.7988313124515116e-05
step: 30, loss: 0.00011849871225422248
step: 40, loss: 8.942127169575542e-05
step: 50, loss: 0.00016671141202095896
step: 60, loss: 0.00017368773114867508
step: 70, loss: 9.001555008580908e-05
step: 80, loss: 0.00015029398491606116
step: 90, loss: 0.0001349527738057077
step: 100, loss: 8.672189142089337e-05
step: 110, loss: 0.002151476452127099
step: 120, loss: 0.000176163885043934
step: 130, loss: 4.5643002522410825e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 140, loss: 6.568845856236294e-05
step: 150, loss: 6.726505671394989e-05
step: 160, loss: 6.633030716329813e-05
step: 170, loss: 0.01797511614859104
step: 180, loss: 6.840439891675487e-05
step: 190, loss: 9.614977170713246e-05
step: 200, loss: 0.00016553897876292467
step: 210, loss: 5.088447142043151e-05
step: 220, loss: 7.38675007596612e-05
step: 230, loss: 0.006454721558839083
/usr/local/lib/python3.7/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=0.9752808988764046, f1=0.968609865470852,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.185359183698893e-05
step: 10, loss: 0.00011827843263745308
step: 20, loss: 5.677532681147568e-05
step: 30, loss: 0.0002583017048891634
step: 40, loss: 4.257368709659204e-05
step: 50, loss: 5.8599758631316945e-05
step: 60, loss: 0.010422628372907639
step: 70, loss: 5.627433711197227e-05
step: 80, loss: 3.6777768400497735e-05
step: 90, loss: 0.0001575181377120316
step: 100, loss: 5.7795979955699295e-05
step: 110, loss: 0.00031604207470081747
step: 120, loss: 0.00010000081965699792
step: 130, loss: 0.00020453099568840116

```

```

step: 140, loss: 6.512695108540356e-05
step: 150, loss: 3.3104908652603626e-05
step: 160, loss: 0.0009749761666171253
step: 170, loss: 8.111106581054628e-05
step: 180, loss: 6.148537067929283e-05
step: 190, loss: 5.6598117225803435e-05
step: 200, loss: 0.0016975807957351208
step: 210, loss: 6.99799056746997e-05
step: 220, loss: 0.001448807306587696
step: 230, loss: 8.186313789337873e-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=0.9787234042553192, f1=0.9744160177975528,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.211498137214221e-05
step: 10, loss: 6.545860378537327e-05
step: 20, loss: 7.230968913063407e-05
step: 30, loss: 0.0010525028919801116
step: 40, loss: 0.00044263555901125073
step: 50, loss: 0.0005010700551792979
step: 60, loss: 9.44023922784254e-05
step: 70, loss: 5.964437877992168e-05
step: 80, loss: 5.038973540649749e-05
step: 90, loss: 3.700912566273473e-05
step: 100, loss: 7.15245187166147e-05
step: 110, loss: 0.00022442407498601824
step: 120, loss: 9.786286682356149e-05
step: 130, loss: 0.02038736455142498
step: 140, loss: 0.0005365342949517071
step: 150, loss: 0.022519174963235855
step: 160, loss: 8.07941542007029e-05
step: 170, loss: 7.099560752976686e-05
step: 180, loss: 0.00016383451293222606
step: 190, loss: 4.081182851223275e-05
step: 200, loss: 9.42320839385502e-05
step: 210, loss: 0.00022905356308910996
step: 220, loss: 0.0005530560156330466
step: 230, loss: 3.975769504904747e-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=0.9799107142857142, f1=0.9744160177975528,
best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.886799433734268e-05
step: 10, loss: 3.2594503863947466e-05
step: 20, loss: 0.016247179359197617
step: 30, loss: 6.983189814491197e-05
step: 40, loss: 0.00042849068995565176
step: 50, loss: 6.692013266729191e-05
step: 60, loss: 0.0001569855521665886
step: 70, loss: 6.575275619979948e-05
step: 80, loss: 6.462108285631984e-05
step: 90, loss: 0.0002283458161400631
step: 100, loss: 0.010723739862442017
step: 110, loss: 0.00015594155411235988
step: 120, loss: 5.8357585658086464e-05
step: 130, loss: 0.00013941098586656153
step: 140, loss: 4.78172478324268e-05
step: 150, loss: 7.937532791402191e-05
step: 160, loss: 3.312736953375861e-05
step: 170, loss: 7.239088154165074e-05
step: 180, loss: 7.918670598883182e-05
step: 190, loss: 0.00032992055639624596
step: 200, loss: 4.6410292270593345e-05
step: 210, loss: 0.02514914609491825
step: 220, loss: 7.010630361037329e-05
step: 230, loss: 2.3822760340408422e-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=0.98, f1=0.9745293466223698, best_f1=0.974472807991121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011320404155412689
step: 10, loss: 6.110388494562358e-05
step: 20, loss: 6.871597724966705e-05
step: 30, loss: 0.00012648256961256266
step: 40, loss: 8.383599197259173e-05
step: 50, loss: 0.00010146848944714293
step: 60, loss: 3.955630018026568e-05
step: 70, loss: 6.237834895728156e-05
step: 80, loss: 0.0001938820059876889
step: 90, loss: 3.254638795624487e-05
step: 100, loss: 0.00014600767462980002
step: 110, loss: 7.35690409783274e-05
step: 120, loss: 9.64850332820788e-05
step: 130, loss: 5.859326847712509e-05
step: 140, loss: 0.012936837039887905
```

```

step: 150, loss: 0.00012434877862688154
step: 160, loss: 9.248300921171904e-05
step: 170, loss: 5.2928586228517815e-05
step: 180, loss: 5.2696945203933865e-05
step: 190, loss: 0.00657310476526618
step: 200, loss: 4.265565439709462e-05
step: 210, loss: 0.033697862178087234
step: 220, loss: 0.011255448684096336
step: 230, loss: 6.819725967943668e-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=0.98, f1=0.9734513274336283, best_f1=0.974472807991121
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 [00:08, 298.07it/s]
load_f1 = 0.9822616407982262
real_f1 = 0.9800884955752212
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, 367.62it/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: 0.7930416464805603
step: 10, loss: 0.40324440598487854
step: 20, loss: 0.48326799273490906
step: 30, loss: 0.4084390699863434
step: 40, loss: 0.30450886487960815
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.1714194267988205
step: 60, loss: 0.20629383623600006
step: 70, loss: 0.21067535877227783
step: 80, loss: 0.08878825604915619
step: 90, loss: 0.1861455738544464
step: 100, loss: 0.21588842570781708
step: 110, loss: 0.0350879468023777
step: 120, loss: 0.020357441157102585
step: 130, loss: 0.014935187995433807
step: 140, loss: 0.16648930311203003
step: 150, loss: 0.037032101303339005
step: 160, loss: 0.14211216568946838
step: 170, loss: 0.10520074516534805
step: 180, loss: 0.1328655630350113
step: 190, loss: 0.028981387615203857
step: 200, loss: 0.08626712113618851
step: 210, loss: 0.06182420253753662
step: 220, loss: 0.20595543086528778
step: 230, loss: 0.07665754109621048

```

```

step: 240, loss: 0.03499038517475128
step: 250, loss: 0.03248974680900574
step: 260, loss: 0.024583080783486366
step: 270, loss: 0.01124020665884018
step: 280, loss: 0.0579272024333477
step: 290, loss: 0.04227447882294655
step: 300, loss: 0.21772488951683044
step: 310, loss: 0.06843589246273041
step: 320, loss: 0.035128433257341385
step: 330, loss: 0.04180010408163071
step: 340, loss: 0.13224682211875916
step: 350, loss: 0.07894118875265121
step: 360, loss: 0.037883199751377106
step: 370, loss: 0.16899225115776062
step: 380, loss: 0.17691783607006073
step: 390, loss: 0.012470640242099762
step: 400, loss: 0.0102731604129076
step: 410, loss: 0.011560061946511269
step: 420, loss: 0.007790336851030588
step: 430, loss: 0.047964148223400116
step: 440, loss: 0.051939595490694046
step: 450, loss: 0.0155989620834589
step: 460, loss: 0.15439093112945557
step: 470, loss: 0.15820148587226868
step: 480, loss: 0.3039890229701996
step: 490, loss: 0.021793464198708534
step: 500, loss: 0.007505025248974562
step: 510, loss: 0.06966952979564667
step: 520, loss: 0.09576024115085602
step: 530, loss: 0.10668329894542694
/usr/local/lib/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=0.9343664539653601, f1=0.9367894497498862,
best_f1=0.9367894497498862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07674366235733032
step: 10, loss: 0.11540735512971878
step: 20, loss: 0.11227165907621384
step: 30, loss: 0.05671825259923935
step: 40, loss: 0.0027511867228895426
step: 50, loss: 0.03063085675239563

```

step: 60, loss: 0.08034927397966385
step: 70, loss: 0.10054640471935272
step: 80, loss: 0.011991029605269432
step: 90, loss: 0.004019488114863634
step: 100, loss: 0.25394919514656067
step: 110, loss: 0.024834588170051575
step: 120, loss: 0.04216376319527626
step: 130, loss: 0.013216427527368069
step: 140, loss: 0.02417156845331192
step: 150, loss: 0.044606953859329224
step: 160, loss: 0.06387896090745926
step: 170, loss: 0.1414390355348587
step: 180, loss: 0.014183438383042812
step: 190, loss: 0.031816404312849045
step: 200, loss: 0.005146542098373175
step: 210, loss: 0.0012688516872003675
step: 220, loss: 0.17252929508686066
step: 230, loss: 0.07928765565156937
step: 240, loss: 0.17560505867004395
step: 250, loss: 0.03621482476592064
step: 260, loss: 0.014218966476619244
step: 270, loss: 0.10888172686100006
step: 280, loss: 0.09989973902702332
step: 290, loss: 0.06829894334077835
step: 300, loss: 0.02624833956360817
step: 310, loss: 0.025564031675457954
step: 320, loss: 0.026148466393351555
step: 330, loss: 0.0027126704808324575
step: 340, loss: 0.005917422939091921
step: 350, loss: 0.03059130162000656
step: 360, loss: 0.00901461485773325
step: 370, loss: 0.0286373533308506
step: 380, loss: 0.09246759861707687
step: 390, loss: 0.03392397612333298
step: 400, loss: 0.023596664890646935
step: 410, loss: 0.0006068867514841259
step: 420, loss: 0.02186218835413456
step: 430, loss: 0.028683260083198547
step: 440, loss: 0.006301736459136009
step: 450, loss: 0.008379832841455936
step: 460, loss: 0.22529298067092896
step: 470, loss: 0.13034208118915558
step: 480, loss: 0.15782247483730316
step: 490, loss: 0.03378789871931076
step: 500, loss: 0.023012908175587654
step: 510, loss: 0.07432754337787628
step: 520, loss: 0.2014603167772293
step: 530, loss: 0.10408148169517517


```

/usr/local/lib/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=0.9469801751959428, f1=0.9450346420323326,
best_f1=0.9450346420323326
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012879610061645508
step: 10, loss: 0.02289699576795101
step: 20, loss: 0.15498311817646027
step: 30, loss: 0.21288281679153442
step: 40, loss: 0.003418492153286934
step: 50, loss: 0.004015639424324036
step: 60, loss: 0.0019602514803409576
step: 70, loss: 0.0034239019732922316
step: 80, loss: 0.000869042647536844
step: 90, loss: 0.0395054817199707
step: 100, loss: 0.05984121933579445
step: 110, loss: 0.025925088673830032
step: 120, loss: 0.018110988661646843
step: 130, loss: 0.06188143789768219
step: 140, loss: 0.020018527284264565
step: 150, loss: 0.01620013453066349
step: 160, loss: 0.0028587786946445704
step: 170, loss: 0.0017269367817789316
step: 180, loss: 0.004825178533792496
step: 190, loss: 0.010248180478811264
step: 200, loss: 0.0046462612226605415
step: 210, loss: 0.05595007166266441
step: 220, loss: 0.009506464004516602
step: 230, loss: 0.011790004558861256
step: 240, loss: 0.011107361875474453
step: 250, loss: 0.003966316115111113
step: 260, loss: 0.0037498795427381992
step: 270, loss: 0.0020367412362247705
step: 280, loss: 0.007507536560297012
step: 290, loss: 0.1076483428478241
step: 300, loss: 0.034504234790802
step: 310, loss: 0.101943239569664
step: 320, loss: 0.08453202247619629
step: 330, loss: 0.0013222872512415051
step: 340, loss: 0.0008972896030172706
step: 350, loss: 0.016691779717803

```

```

step: 360, loss: 0.024207020178437233
step: 370, loss: 0.0008331204880960286
step: 380, loss: 0.1004352793097496
step: 390, loss: 0.013023675419390202
step: 400, loss: 0.02115788869559765
step: 410, loss: 0.0025678740348666906
step: 420, loss: 0.01999283768236637
step: 430, loss: 0.05455530807375908
step: 440, loss: 0.019053412601351738
step: 450, loss: 0.017790669575333595
step: 460, loss: 0.032145772129297256
step: 470, loss: 0.0023104106076061726
step: 480, loss: 0.002014412544667721
step: 490, loss: 0.005619964096695185
step: 500, loss: 0.026704175397753716
step: 510, loss: 0.0040424661710858345
step: 520, loss: 0.009149137884378433
step: 530, loss: 0.0063608610071241856
/usr/local/lib/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=0.9491525423728814, f1=0.9476102941176471,
best_f1=0.9476102941176471
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016177416546270251
step: 10, loss: 0.005767556838691235
step: 20, loss: 0.030350888147950172
step: 30, loss: 0.0073120673187077045
step: 40, loss: 0.0030703400261700153
step: 50, loss: 0.012640012428164482
step: 60, loss: 0.00028160391957499087
step: 70, loss: 0.0043225460685789585
step: 80, loss: 0.050513193011283875
step: 90, loss: 0.0008925010333769023
step: 100, loss: 0.00046466238563880324
step: 110, loss: 0.0001414694997947663
step: 120, loss: 0.016696225851774216
step: 130, loss: 0.008236595429480076
step: 140, loss: 0.03292686119675636
step: 150, loss: 0.0018242100486531854
step: 160, loss: 0.0010824992787092924
step: 170, loss: 0.004345269408077002

```

```

step: 180, loss: 0.00393136078491807
step: 190, loss: 0.00032843402004800737
step: 200, loss: 0.0010195523500442505
step: 210, loss: 0.0037020305171608925
step: 220, loss: 0.015344671905040741
step: 230, loss: 0.21454519033432007
step: 240, loss: 0.015433385036885738
step: 250, loss: 0.006531815975904465
step: 260, loss: 0.08158575743436813
step: 270, loss: 0.09151991456747055
step: 280, loss: 0.008758791722357273
step: 290, loss: 0.02985503152012825
step: 300, loss: 0.03095942921936512
step: 310, loss: 0.007072695065289736
step: 320, loss: 0.009643112309277058
step: 330, loss: 0.020513448864221573
step: 340, loss: 0.0014370434219017625
step: 350, loss: 0.0007559806690551341
step: 360, loss: 0.000652490125503391
step: 370, loss: 0.03477563336491585
step: 380, loss: 0.0007248759502544999
step: 390, loss: 0.013617626391351223
step: 400, loss: 0.01605100743472576
step: 410, loss: 0.01924118772149086
step: 420, loss: 0.020474303513765335
step: 430, loss: 0.00379791297018528
step: 440, loss: 0.018317798152565956
step: 450, loss: 0.013307997956871986
step: 460, loss: 0.00031985825626179576
step: 470, loss: 0.005859612487256527
step: 480, loss: 0.0007306317565962672
step: 490, loss: 0.013927650637924671
step: 500, loss: 0.004813822917640209
step: 510, loss: 0.015029548667371273
step: 520, loss: 0.004618310835212469
step: 530, loss: 0.030392179265618324
/usr/local/lib/python3.7/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=0.9489795918367347, f1=0.9471715755025713,
best_f1=0.9476102941176471
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0026602379512041807
step: 10, loss: 0.0010879928013309836
step: 20, loss: 0.09231235831975937
step: 30, loss: 0.005431246943771839

```

step: 40, loss: 0.1279861032962799
step: 50, loss: 0.03384154662489891
step: 60, loss: 0.006249387748539448
step: 70, loss: 0.0003643868549261242
step: 80, loss: 0.0008044351707212627
step: 90, loss: 0.01923176273703575
step: 100, loss: 0.0003757632221095264
step: 110, loss: 0.000583229586482048
step: 120, loss: 0.011901509948074818
step: 130, loss: 0.0007711219368502498
step: 140, loss: 0.0006530453683808446
step: 150, loss: 0.00028201795066706836
step: 160, loss: 0.008621438406407833
step: 170, loss: 0.02160695567727089
step: 180, loss: 0.0012959151063114405
step: 190, loss: 0.00015859759878367186
step: 200, loss: 0.00029127084417268634
step: 210, loss: 0.00028061599005013704
step: 220, loss: 9.030865476233885e-05
step: 230, loss: 0.00012286777200642973
step: 240, loss: 0.0045101577416062355
step: 250, loss: 0.0002684287610463798
step: 260, loss: 0.016499273478984833
step: 270, loss: 0.04419877007603645
step: 280, loss: 0.013218099251389503
step: 290, loss: 0.00014128514158073813
step: 300, loss: 0.0008880353416316211
step: 310, loss: 0.0003597811155486852
step: 320, loss: 0.0021133781410753727
step: 330, loss: 0.0009646312682889402
step: 340, loss: 0.002783876610919833
step: 350, loss: 0.08201944082975388
step: 360, loss: 0.28932610154151917
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 370, loss: 0.0010038010077551007
step: 380, loss: 0.019703498110175133
step: 390, loss: 0.022238096222281456
step: 400, loss: 0.0023922869004309177
step: 410, loss: 0.00045828463044017553
step: 420, loss: 0.003939674701541662
step: 430, loss: 0.0032915095798671246
step: 440, loss: 0.00041138488450087607
step: 450, loss: 0.012877121567726135
step: 460, loss: 0.0008527233148925006
step: 470, loss: 0.0075570871122181416
step: 480, loss: 0.030232157558202744
step: 490, loss: 0.005837631411850452
step: 500, loss: 0.02488095499575138

```

step: 510, loss: 0.026887821033596992
step: 520, loss: 0.00041398583562113345
step: 530, loss: 0.0019511731807142496
/usr/local/lib/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=0.9505773672055428, f1=0.9488243430152145,
best_f1=0.9488243430152145
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005541897262446582
step: 10, loss: 0.0003962784830946475
step: 20, loss: 0.00016273451910819858
step: 30, loss: 0.0003211887087672949
step: 40, loss: 0.00026937059010379016
step: 50, loss: 0.0001690225617494434
step: 60, loss: 0.0004282171721570194
step: 70, loss: 0.014181781560182571
step: 80, loss: 0.00013401004252955317
step: 90, loss: 0.0002664563653524965
step: 100, loss: 0.13741056621074677
step: 110, loss: 0.012040212750434875
step: 120, loss: 0.000701627810485661
step: 130, loss: 0.0011183316819369793
step: 140, loss: 0.0028151029255241156
step: 150, loss: 0.0013106907717883587
step: 160, loss: 0.0005170927615836263
step: 170, loss: 0.0006772616179659963
step: 180, loss: 0.00017464760458096862
step: 190, loss: 0.0010421971092000604
step: 200, loss: 8.719330071471632e-05
step: 210, loss: 0.0001101478555938229
step: 220, loss: 0.010044550523161888
step: 230, loss: 9.604065417079255e-05
step: 240, loss: 0.0006850005011074245
step: 250, loss: 0.006776873487979174
step: 260, loss: 0.0006292075850069523
step: 270, loss: 0.0004052390868309885
step: 280, loss: 0.0035798356402665377
step: 290, loss: 0.031632017344236374
step: 300, loss: 0.006087673362344503
step: 310, loss: 0.00043117470340803266
step: 320, loss: 0.05592546984553337

```

```

step: 330, loss: 0.183585986495018
step: 340, loss: 0.027087723836302757
step: 350, loss: 0.08071383833885193
step: 360, loss: 0.005539397243410349
step: 370, loss: 0.001579401665367186
step: 380, loss: 0.0097163375467062
step: 390, loss: 0.002498153829947114
step: 400, loss: 0.000336087221512571
step: 410, loss: 0.0001306883932556957
step: 420, loss: 0.007074104622006416
step: 430, loss: 0.0009661615476943552
step: 440, loss: 0.0008699835161678493
step: 450, loss: 0.002150978660210967
step: 460, loss: 0.004276682157069445
step: 470, loss: 0.005635683890432119
step: 480, loss: 0.01166726928204298
step: 490, loss: 0.0001437178289052099
step: 500, loss: 0.00027440168196335435
step: 510, loss: 0.00021176901645958424
step: 520, loss: 0.0036361522506922483
step: 530, loss: 0.0042176516726613045
/usr/local/lib/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=0.9525581395348838, f1=0.9493258949325895,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000723455857951194
step: 10, loss: 0.0011501952540129423
step: 20, loss: 0.004439800512045622
step: 30, loss: 0.006434316281229258
step: 40, loss: 0.00047573892516084015
step: 50, loss: 0.0025031077675521374
step: 60, loss: 0.04887208342552185
step: 70, loss: 0.000245341012487188
step: 80, loss: 0.0014413039898499846
step: 90, loss: 0.0001751614618115127
step: 100, loss: 0.005794224329292774
step: 110, loss: 0.0012495029950514436
step: 120, loss: 0.0006160488701425493
step: 130, loss: 0.00011880915553774685
step: 140, loss: 0.0038871830329298973

```

```

step: 150, loss: 8.573894592700526e-05
step: 160, loss: 9.054921247297898e-05
step: 170, loss: 0.0004286832408979535
step: 180, loss: 6.437363481381908e-05
step: 190, loss: 0.00015416447422467172
step: 200, loss: 0.00012237968621775508
step: 210, loss: 9.346046863356605e-05
step: 220, loss: 0.0021267449483275414
step: 230, loss: 9.96955786831677e-05
step: 240, loss: 0.004950052592903376
step: 250, loss: 0.0002030088216997683
step: 260, loss: 0.0006416531396098435
step: 270, loss: 8.179629367077723e-05
step: 280, loss: 0.009672190994024277
step: 290, loss: 0.00127050606533885
step: 300, loss: 0.00023088361194822937
step: 310, loss: 0.00020417748601175845
step: 320, loss: 0.03393168747425079
step: 330, loss: 0.00011420324153732508
step: 340, loss: 0.0038584682624787092
step: 350, loss: 8.991920913103968e-05
step: 360, loss: 0.0003767232410609722
step: 370, loss: 7.459419430233538e-05
step: 380, loss: 0.0028905083891004324
step: 390, loss: 5.3929343266645446e-05
step: 400, loss: 8.308229735121131e-05
step: 410, loss: 0.0020769776310771704
step: 420, loss: 0.0016246411250904202
step: 430, loss: 3.5697990824701265e-05
step: 440, loss: 0.00012525040074251592
step: 450, loss: 0.0002004305279115215
step: 460, loss: 0.0034026445355266333
step: 470, loss: 0.0018245266983285546
step: 480, loss: 0.0007938713533803821
step: 490, loss: 0.00030610724934376776
step: 500, loss: 6.952000694582239e-05
step: 510, loss: 0.007181545253843069
step: 520, loss: 0.00023406470427289605
step: 530, loss: 0.00018374391947872937
/usr/local/lib/python3.7/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=0.9486461251167133, f1=0.9384109073812883,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006542739574797451

```

step: 10, loss: 0.0004375274002086371
step: 20, loss: 0.024451564997434616
step: 30, loss: 0.008609993383288383
step: 40, loss: 0.00047621215344406664
step: 50, loss: 0.00042272292193956673
step: 60, loss: 0.00019354105461388826
step: 70, loss: 6.390440830728039e-05
step: 80, loss: 6.304528506007046e-05
step: 90, loss: 0.0004904110101051629
step: 100, loss: 0.00019987062842119485
step: 110, loss: 0.00020741720800288022
step: 120, loss: 8.544366573914886e-05
step: 130, loss: 0.00018119426385965198
step: 140, loss: 2.4839744583005086e-05
step: 150, loss: 8.950305345933884e-05
step: 160, loss: 0.00016011040133889765
step: 170, loss: 3.4178545320173725e-05
step: 180, loss: 0.00021991468383930624
step: 190, loss: 0.00015567276568617672
step: 200, loss: 0.0010669452603906393
step: 210, loss: 0.0021954786498099566
step: 220, loss: 0.0004553399048745632
step: 230, loss: 0.0006805752636864781
step: 240, loss: 4.796373104909435e-05
step: 250, loss: 8.163481106748804e-05
step: 260, loss: 0.0006903646280989051
step: 270, loss: 2.9193897717050277e-05
step: 280, loss: 0.0006106987129896879
step: 290, loss: 0.0006080001476220787
step: 300, loss: 4.246254684403539e-05
step: 310, loss: 0.013801982626318932
step: 320, loss: 0.00022309496125672013
step: 330, loss: 0.023638665676116943
step: 340, loss: 6.385389133356512e-05
step: 350, loss: 0.1576295644044876
step: 360, loss: 0.0015847994945943356
step: 370, loss: 0.00040377379627898335
step: 380, loss: 0.029918506741523743
step: 390, loss: 0.0002747494145296514
step: 400, loss: 0.032679442316293716
step: 410, loss: 0.00015796773368492723
step: 420, loss: 3.584736259654164e-05
step: 430, loss: 0.004474022891372442
step: 440, loss: 0.00037439874722622335
step: 450, loss: 0.0010073587764054537
step: 460, loss: 0.0008088565082289279
step: 470, loss: 0.00011191431985935196
step: 480, loss: 0.0339314341545105


```

step: 490, loss: 0.00014863903925288469
step: 500, loss: 0.0009033208480104804
step: 510, loss: 5.743870497099124e-05
step: 520, loss: 8.416769560426474e-05
step: 530, loss: 0.004746378865092993
/usr/local/lib/python3.7/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=0.9476145930776427, f1=0.9449411764705883,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002616901183500886
step: 10, loss: 0.0003390830825082958
step: 20, loss: 0.00047544564586132765
step: 30, loss: 4.863475260208361e-05
step: 40, loss: 0.0025949287228286266
step: 50, loss: 0.00020396229228936136
step: 60, loss: 0.0003033480606973171
step: 70, loss: 0.02986013889312744
step: 80, loss: 0.0011038183001801372
step: 90, loss: 0.0003984592040069401
step: 100, loss: 6.648324779234827e-05
step: 110, loss: 0.00010452845162944868
step: 120, loss: 3.96386458305642e-05
step: 130, loss: 5.13412851432804e-05
step: 140, loss: 0.02539210021495819
step: 150, loss: 6.843356823083013e-05
step: 160, loss: 0.00012501196761149913
step: 170, loss: 6.743835547240451e-05
step: 180, loss: 0.00596441188827157
step: 190, loss: 0.0002170894294977188
step: 200, loss: 0.0005535379168577492
step: 210, loss: 0.0006094294367358088
step: 220, loss: 0.0007155233179219067
step: 230, loss: 0.0002847848227247596
step: 240, loss: 0.00021170520631130785
step: 250, loss: 9.744147973833606e-05
step: 260, loss: 0.001131475786678493
step: 270, loss: 0.0010014197323471308
step: 280, loss: 9.04326225281693e-05
step: 290, loss: 3.1786410545464605e-05
step: 300, loss: 4.5303771912585944e-05
step: 310, loss: 2.8910986657137983e-05
step: 320, loss: 0.00010394472337793559
step: 330, loss: 4.439811164047569e-05
step: 340, loss: 0.00017284831847064197

```

```

step: 350, loss: 3.104164352407679e-05
step: 360, loss: 0.01283032912760973
step: 370, loss: 0.00028258253587409854
step: 380, loss: 3.2821582863107324e-05
step: 390, loss: 5.4438351071439683e-05
step: 400, loss: 0.00015328491281252354
step: 410, loss: 7.235897646751255e-05
step: 420, loss: 0.0003859831194858998
step: 430, loss: 3.561990524758585e-05
step: 440, loss: 4.7867528337519616e-05
step: 450, loss: 4.8298665205948055e-05
step: 460, loss: 0.004432902671396732
step: 470, loss: 2.8836708224844187e-05
step: 480, loss: 5.60810549359303e-05
step: 490, loss: 4.620895197149366e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 500, loss: 0.016293134540319443
step: 510, loss: 0.0004216250090394169
step: 520, loss: 3.705979906953871e-05
step: 530, loss: 6.416615360649303e-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=0.9466357308584686, f1=0.9442379182156133,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022192718461155891
step: 10, loss: 1.8693235688260756e-05
step: 20, loss: 6.411707727238536e-05
step: 30, loss: 4.211488339933567e-05
step: 40, loss: 5.2439536375459284e-05
step: 50, loss: 0.042454320937395096
step: 60, loss: 0.00012141295155743137
step: 70, loss: 0.00010477491014171392
step: 80, loss: 0.0006935730343684554
step: 90, loss: 0.000562203349545598
step: 100, loss: 0.00017822410154622048
step: 110, loss: 0.00014687373186461627
step: 120, loss: 0.0001142267428804189
step: 130, loss: 2.58267446042737e-05
step: 140, loss: 0.0009252906893379986
step: 150, loss: 0.00013447956007439643
step: 160, loss: 5.1428069127723575e-05
step: 170, loss: 0.00022484238434117287
step: 180, loss: 0.19803419709205627
step: 190, loss: 0.001075355219654739

```

```

step: 200, loss: 0.003061671508476138
step: 210, loss: 0.0048280940391123295
step: 220, loss: 0.0005121464491821826
step: 230, loss: 0.002166198333725333
step: 240, loss: 6.860142457298934e-05
step: 250, loss: 0.0032371904235333204
step: 260, loss: 0.022529391571879387
step: 270, loss: 0.00011933455243706703
step: 280, loss: 9.197094914270565e-05
step: 290, loss: 0.00011888073640875518
step: 300, loss: 0.0005411054589785635
step: 310, loss: 6.335984653560445e-05
step: 320, loss: 4.1378531022928655e-05
step: 330, loss: 5.957460598438047e-05
step: 340, loss: 0.002185130724683404
step: 350, loss: 4.403237107908353e-05
step: 360, loss: 0.00021325211855582893
step: 370, loss: 0.03095567412674427
step: 380, loss: 8.894364873412997e-05
step: 390, loss: 0.0007237648824229836
step: 400, loss: 0.00024630961706861854
step: 410, loss: 7.802501932019368e-05
step: 420, loss: 0.009080719202756882
step: 430, loss: 3.589796688174829e-05
step: 440, loss: 3.675051993923262e-05
step: 450, loss: 0.0004949792055413127
step: 460, loss: 0.0030849548056721687
step: 470, loss: 7.087855919962749e-05
step: 480, loss: 8.216589048970491e-05
step: 490, loss: 0.06016984581947327
step: 500, loss: 0.02193278633058071
step: 510, loss: 0.004522424191236496
step: 520, loss: 0.0007161421235650778
step: 530, loss: 3.685596675495617e-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=0.9491682070240297, f1=0.9473197781885397,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.5494281948776916e-05
step: 10, loss: 0.0002154832473024726
step: 20, loss: 2.3960223188623786e-05
step: 30, loss: 0.00039000908145681024
step: 40, loss: 0.03468892350792885
step: 50, loss: 0.055259667336940765

```

step: 60, loss: 5.164767935639247e-05
step: 70, loss: 2.558085361670237e-05
step: 80, loss: 0.0008315882878378034
step: 90, loss: 0.015119067393243313
step: 100, loss: 0.0006239045760594308
step: 110, loss: 7.864144572522491e-05
step: 120, loss: 0.0017260705353692174
step: 130, loss: 2.3032978788251057e-05
step: 140, loss: 1.9434512068983167e-05
step: 150, loss: 4.6358731196960434e-05
step: 160, loss: 0.0002598315477371216
step: 170, loss: 9.879298158921301e-05
step: 180, loss: 2.402757127129007e-05
step: 190, loss: 0.00012172774586360902
step: 200, loss: 0.000345858366927132
step: 210, loss: 3.6759105569217354e-05
step: 220, loss: 6.855368701508269e-05
step: 230, loss: 3.891803135047667e-05
step: 240, loss: 5.233427509665489e-05
step: 250, loss: 4.3950883991783485e-05
step: 260, loss: 2.5372366508236155e-05
step: 270, loss: 0.00018445326713845134
step: 280, loss: 3.130643744952977e-05
step: 290, loss: 0.0022461258340626955
step: 300, loss: 0.0008986227330751717
step: 310, loss: 0.000570675649214536
step: 320, loss: 0.028262969106435776
step: 330, loss: 1.9877776139765047e-05
step: 340, loss: 2.028756898653228e-05
step: 350, loss: 0.00017615652177482843
step: 360, loss: 0.00014427224232349545
step: 370, loss: 1.8860793716157787e-05
step: 380, loss: 5.095290907775052e-05
step: 390, loss: 0.00028748682234436274
step: 400, loss: 3.0287821573438123e-05
step: 410, loss: 2.4567340005887672e-05
step: 420, loss: 4.285563045414165e-05
step: 430, loss: 0.006633236538618803
step: 440, loss: 4.197216185275465e-05
step: 450, loss: 4.565685230772942e-05
step: 460, loss: 0.005612242966890335
step: 470, loss: 8.568657358409837e-05
step: 480, loss: 2.278718602610752e-05
step: 490, loss: 0.0015418013790622354
step: 500, loss: 2.4347884391318075e-05
step: 510, loss: 2.632589894346893e-05
step: 520, loss: 1.888690167106688e-05
step: 530, loss: 3.201952495146543e-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=0.9517564402810305, f1=0.9424528301886792,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.9786328013869934e-05
step: 10, loss: 3.537752854754217e-05
step: 20, loss: 0.02417575940489769
step: 30, loss: 2.0600489733624272e-05
step: 40, loss: 6.620090425712988e-05
step: 50, loss: 0.002753038424998522
step: 60, loss: 4.0152284782379866e-05
step: 70, loss: 4.783396434504539e-05
step: 80, loss: 7.379520684480667e-05
step: 90, loss: 0.0012598622124642134
step: 100, loss: 6.603399378946051e-05
step: 110, loss: 0.04707358777523041
step: 120, loss: 8.716641605133191e-05
step: 130, loss: 3.3111853554146364e-05
step: 140, loss: 0.00015759962843731046
step: 150, loss: 3.7441706808749586e-05
step: 160, loss: 0.0020245276391506195
step: 170, loss: 2.4343866243725643e-05
step: 180, loss: 0.00011055759387090802
step: 190, loss: 6.16672114119865e-05
step: 200, loss: 2.5189674488501623e-05
step: 210, loss: 0.0002592363744042814
step: 220, loss: 2.568144554970786e-05
step: 230, loss: 4.5391032472252846e-05
step: 240, loss: 7.724767056060955e-05
step: 250, loss: 0.00011086879385402426
step: 260, loss: 4.1944840631913394e-05
step: 270, loss: 2.7588283046497963e-05
step: 280, loss: 3.77360156562645e-05
step: 290, loss: 0.0010435734875500202
step: 300, loss: 0.018940165638923645
step: 310, loss: 5.4638294386677444e-05
step: 320, loss: 2.5811805244302377e-05
step: 330, loss: 1.8704386093304493e-05
step: 340, loss: 2.6667923521017656e-05
step: 350, loss: 0.02142123319208622
step: 360, loss: 0.00021538529836107045
step: 370, loss: 6.200054485816509e-05
step: 380, loss: 3.485520574031398e-05
step: 390, loss: 2.0790463167941198e-05

```

```

step: 400, loss: 2.7148660592501983e-05
step: 410, loss: 0.0045350524596869946
step: 420, loss: 4.4813506974605843e-05
step: 430, loss: 2.057061647064984e-05
step: 440, loss: 2.754598608589731e-05
step: 450, loss: 0.00013994517212267965
step: 460, loss: 4.061897925566882e-05
step: 470, loss: 3.270254092058167e-05
step: 480, loss: 2.9533302949857898e-05
step: 490, loss: 6.832463986938819e-05
step: 500, loss: 6.292843318078667e-05
step: 510, loss: 2.8240739993634634e-05
step: 520, loss: 0.018616050481796265
step: 530, loss: 0.0003614122106228024
/usr/local/lib/python3.7/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=0.9482109227871939, f1=0.9408983451536642,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.8436613976955414e-05
step: 10, loss: 1.7281305190408602e-05
step: 20, loss: 0.023425761610269547
step: 30, loss: 0.0017174248350784183
step: 40, loss: 1.6499052435392514e-05
step: 50, loss: 3.0317431082949042e-05
step: 60, loss: 0.010675848461687565
step: 70, loss: 1.6044563381001353e-05
step: 80, loss: 1.7944366845767945e-05
step: 90, loss: 0.0028179020155221224
step: 100, loss: 2.496220804459881e-05
step: 110, loss: 0.0006011886871419847
step: 120, loss: 3.348100290168077e-05
step: 130, loss: 3.324984209029935e-05
step: 140, loss: 0.0005191860836930573
step: 150, loss: 4.3228876165812835e-05
step: 160, loss: 2.3144424631027505e-05
step: 170, loss: 0.0005466933362185955
step: 180, loss: 2.0119752662139945e-05
step: 190, loss: 2.9462016755132936e-05
step: 200, loss: 4.438131145434454e-05
step: 210, loss: 0.00011811593867605552
step: 220, loss: 2.302500797668472e-05
step: 230, loss: 0.003503895830363035
step: 240, loss: 6.337927334243432e-05
step: 250, loss: 0.06293534487485886

```

```

step: 260, loss: 8.917348168324679e-05
step: 270, loss: 3.119815664831549e-05
step: 280, loss: 1.8524950064602308e-05
step: 290, loss: 2.695904186111875e-05
step: 300, loss: 3.631869185483083e-05
step: 310, loss: 2.4167356968973763e-05
step: 320, loss: 0.0003245561383664608
step: 330, loss: 2.1684336388716474e-05
step: 340, loss: 3.971907062805258e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 350, loss: 0.015249921008944511
step: 360, loss: 3.189102790202014e-05
step: 370, loss: 2.1430319975479506e-05
step: 380, loss: 4.47022684966214e-05
step: 390, loss: 2.358408528380096e-05
step: 400, loss: 2.9574372092611156e-05
step: 410, loss: 7.942309457575902e-05
step: 420, loss: 1.7765631127986126e-05
step: 430, loss: 0.0008103916188701987
step: 440, loss: 5.317298200679943e-05
step: 450, loss: 0.0003186181711498648
step: 460, loss: 1.7258964362554252e-05
step: 470, loss: 0.024480674415826797
step: 480, loss: 2.6619298296282068e-05
step: 490, loss: 2.8563585146912374e-05
step: 500, loss: 4.838174572796561e-05
step: 510, loss: 1.6491609130753204e-05
step: 520, loss: 1.5571475159958936e-05
step: 530, loss: 1.9709950720425695e-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=0.9523364485981308, f1=0.943449575871819,
best_f1=0.9493258949325895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.545509753166698e-05
step: 10, loss: 0.0002673481358215213
step: 20, loss: 1.405908460583305e-05
step: 30, loss: 4.134585105930455e-05
step: 40, loss: 0.007289331406354904
step: 50, loss: 2.3740129108773544e-05
step: 60, loss: 2.4004992155823857e-05
step: 70, loss: 1.585458085173741e-05
step: 80, loss: 4.4299511500867084e-05
step: 90, loss: 1.2669587704294827e-05
step: 100, loss: 5.165064430912025e-05

```

```
step: 110, loss: 2.202994073741138e-05
step: 120, loss: 3.042478419956751e-05
step: 130, loss: 8.7650420027785e-05
step: 140, loss: 2.519323788874317e-05
step: 150, loss: 0.00015549530507996678
step: 160, loss: 2.749318264250178e-05
step: 170, loss: 4.6161912905517966e-05
step: 180, loss: 4.0354658267460763e-05
step: 190, loss: 2.1844523871550336e-05
step: 200, loss: 0.00025878087035380304
step: 210, loss: 7.179641397669911e-05
step: 220, loss: 1.605949000804685e-05
step: 230, loss: 2.6053627152577974e-05
step: 240, loss: 1.697204788797535e-05
step: 250, loss: 4.722942685475573e-05
step: 260, loss: 1.1414187611080706e-05
step: 270, loss: 4.2358136852271855e-05
step: 280, loss: 2.043226959358435e-05
step: 290, loss: 1.787354631233029e-05
step: 300, loss: 1.837984200392384e-05
step: 310, loss: 1.5474595784326084e-05
step: 320, loss: 2.0097602828172967e-05
step: 330, loss: 4.9249538278672844e-05
step: 340, loss: 2.6020334189524874e-05
step: 350, loss: 6.219396163942292e-05
step: 360, loss: 0.0011251389514654875
step: 370, loss: 1.5377785530290566e-05
step: 380, loss: 1.4453904441324994e-05
step: 390, loss: 1.7430078514735214e-05
step: 400, loss: 2.0324734578025527e-05
step: 410, loss: 1.413358586432878e-05
step: 420, loss: 2.356079130549915e-05
step: 430, loss: 1.7612894225749187e-05
step: 440, loss: 1.1999052730971016e-05
step: 450, loss: 1.414848338754382e-05
step: 460, loss: 2.1766285499325022e-05
step: 470, loss: 1.852131390478462e-05
step: 480, loss: 1.5236237231874838e-05
step: 490, loss: 1.6309022612404078e-05
step: 500, loss: 2.3538919776910916e-05
step: 510, loss: 2.4216642486862838e-05
step: 520, loss: 2.6411253202240914e-05
step: 530, loss: 1.5392719433293678e-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=0.9514472455648927, f1=0.9431123648330982,
best_f1=0.9493258949325895
```


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

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

step: 0, loss: 2.319586565135978e-05
step: 10, loss: 2.163975113944616e-05
step: 20, loss: 1.6934895029407926e-05
step: 30, loss: 2.8760408895323053e-05
step: 40, loss: 1.1365767022653017e-05
step: 50, loss: 1.4737015590071678e-05
step: 60, loss: 2.1414718503365293e-05
step: 70, loss: 1.482271636632504e-05
step: 80, loss: 1.8477154299034737e-05
step: 90, loss: 1.5519310181844048e-05
step: 100, loss: 1.7508484233985655e-05
step: 110, loss: 0.0096842460334301
step: 120, loss: 3.4928212699014693e-05
step: 130, loss: 1.1183243259438314e-05
step: 140, loss: 1.3407146980171092e-05
step: 150, loss: 7.586782885482535e-05
step: 160, loss: 3.158022082061507e-05
step: 170, loss: 1.613393033039756e-05
step: 180, loss: 1.2818598406738602e-05
step: 190, loss: 7.953681779326871e-05
step: 200, loss: 0.0002707427483983338
step: 210, loss: 1.3343809769139625e-05
step: 220, loss: 1.641320704948157e-05
step: 230, loss: 3.7270983739290386e-05
step: 240, loss: 3.055330671486445e-05
step: 250, loss: 8.434631308773533e-05
step: 260, loss: 1.893505032057874e-05
step: 270, loss: 4.217727473587729e-05
step: 280, loss: 1.6093052181531675e-05
step: 290, loss: 5.401793896453455e-05
step: 300, loss: 0.026076624169945717
step: 310, loss: 4.406586231198162e-05
step: 320, loss: 2.513744584575761e-05
step: 330, loss: 1.581363358127419e-05
step: 340, loss: 3.083646879531443e-05
step: 350, loss: 2.36242212849902e-05
step: 360, loss: 3.4611370210768655e-05
step: 370, loss: 1.4282606571214274e-05
step: 380, loss: 5.158774001756683e-05
step: 390, loss: 1.3205967661633622e-05
step: 400, loss: 1.4170830581861082e-05
step: 410, loss: 2.9875187465222552e-05
step: 420, loss: 2.210516140621621e-05
step: 430, loss: 1.871168387879152e-05
step: 440, loss: 0.0007118579815141857

```

step: 450, loss: 7.325961632886901e-05
step: 460, loss: 1.0512688277231064e-05
step: 470, loss: 1.2557826266856864e-05
step: 480, loss: 1.2840911040257197e-05
step: 490, loss: 3.175881283823401e-05
step: 500, loss: 2.266033880005125e-05
step: 510, loss: 2.112910988216754e-05
step: 520, loss: 4.068662383360788e-05
step: 530, loss: 1.8696804545470513e-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=0.9508656995788489, f1=0.9424528301886792,
best_f1=0.9493258949325895
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 [00:16, 343.89it/s]
load_f1 = 0.9535315985130112
real_f1 = 0.9516728624535317
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 [00:11, 374.18it/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 (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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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=877ef70a8686a085ad4f1f9be30e216438294d9ec3a7ff7a7fc78d19e9577588
  Stored in directory: /tmp/pip-ephem-wheel-cache-3fallyj6/wheels/11/95/ba/2c36c
c368c0bd339b44a791c2c1881a1fb714b78c29a4cb8f5
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: 0.8541147112846375
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=0.4, f1=0.2, best_f1=0.2
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37404051423072815
/usr/local/lib/python3.7/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=0.35294117647058826, f1=0.23529411764705882, best_f1=0.2
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.31358271837234497
/usr/local/lib/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=0.5185185185185186, f1=0.2608695652173913,
best_f1=0.2608695652173913
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3389081358909607
/usr/local/lib/python3.7/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=0.4705882352941177, f1=0.40816326530612246,
best_f1=0.2608695652173913
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21176397800445557
/usr/local/lib/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=0.5333333333333333, f1=0.5128205128205129,
best_f1=0.5128205128205129
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2230241298675537
/usr/local/lib/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=0.5789473684210527, f1=0.4210526315789474,
best_f1=0.4210526315789474
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2237672358751297
/usr/local/lib/python3.7/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=0.5454545454545455, f1=0.43478260869565216,
best_f1=0.4210526315789474
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2692681550979614
/usr/local/lib/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=0.5833333333333334, f1=0.4444444444444445,
best_f1=0.4444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1684747338294983
/usr/local/lib/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=0.6666666666666666, f1=0.4615384615384615,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2486249953508377
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.4615384615384615,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20001138746738434
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.4137931034482759,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2167518436908722
/usr/local/lib/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=0.7200000000000001, f1=0.4615384615384615,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13315074145793915
/usr/local/lib/python3.7/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=0.7142857142857143, f1=0.42857142857142855,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16048984229564667
/usr/local/lib/python3.7/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=0.7142857142857143, f1=0.42857142857142855,
best_f1=0.4615384615384615
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2467150092124939
/usr/local/lib/python3.7/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=0.7142857142857143, f1=0.42857142857142855,
best_f1=0.4615384615384615
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, 138591.74it/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 = 0.6923076923076924
real_f1 = 0.6206896551724138
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, 268.21it/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: 0.8071557283401489

step: 10, loss: 0.46394577622413635

step: 20, loss: 0.5610611438751221

step: 30, loss: 0.3357824981212616

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

step: 40, loss: 0.14504103362560272

step: 50, loss: 0.07621487975120544

step: 60, loss: 0.11225621402263641

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0

step: 70, loss: 0.19762153923511505

step: 80, loss: 0.11661432683467865

step: 90, loss: 0.014861596748232841

step: 100, loss: 0.18968328833580017

step: 110, loss: 0.06026739254593849

step: 120, loss: 0.014379756525158882

step: 130, loss: 0.07192233949899673

step: 140, loss: 0.014605163596570492

step: 150, loss: 0.11468993127346039

step: 160, loss: 0.21647459268569946

step: 170, loss: 0.004829712212085724

step: 180, loss: 0.025219671428203583

step: 190, loss: 0.010613963007926941

step: 200, loss: 0.003948111552745104

step: 210, loss: 0.004600050859153271

step: 220, loss: 0.0024436258245259523

step: 230, loss: 0.003788085887208581

/usr/local/lib/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=0.9898762654668166, f1=0.9898534385569334,

best_f1=0.9898534385569334

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

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

step: 0, loss: 0.004463660065084696

step: 10, loss: 0.0008192856912501156

step: 20, loss: 0.0025553491432219744

step: 30, loss: 0.01026548258960247

```

step: 40, loss: 0.0021680269856005907
step: 50, loss: 0.002079853555187583
step: 60, loss: 0.003917039837688208
step: 70, loss: 0.07056570053100586
step: 80, loss: 0.002746753627434373
step: 90, loss: 0.1880536824464798
step: 100, loss: 0.0039135063998401165
step: 110, loss: 0.004968674853444099
step: 120, loss: 0.0008329215925186872
step: 130, loss: 0.0020051731262356043
step: 140, loss: 0.2622734308242798
step: 150, loss: 0.01651410199701786
step: 160, loss: 0.005323197226971388
step: 170, loss: 0.012170786038041115
step: 180, loss: 0.0066515542566776276
step: 190, loss: 0.11959492415189743
step: 200, loss: 0.0077310833148658276
step: 210, loss: 0.03799992427229881
step: 220, loss: 0.0008096006931737065
step: 230, loss: 0.04075711593031883
/usr/local/lib/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=0.9910112359550561, f1=0.9820224719101124,
best_f1=0.9820224719101124
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13847078382968903
step: 10, loss: 0.003067182842642069
step: 20, loss: 0.004843618720769882
step: 30, loss: 0.018016276881098747
step: 40, loss: 0.023712269961833954
step: 50, loss: 0.009109861217439175
step: 60, loss: 0.005207509268075228
step: 70, loss: 0.0007883720099925995
step: 80, loss: 0.000805117073468864
step: 90, loss: 0.001337965251877904
step: 100, loss: 0.0005348594859242439
step: 110, loss: 0.000938068435061723
step: 120, loss: 0.0008128471672534943
step: 130, loss: 0.00027900381246581674
step: 140, loss: 0.002920794766396284
step: 150, loss: 0.0008614562102593482

```

```

step: 160, loss: 0.0027417002711445093
step: 170, loss: 0.0017283889465034008
step: 180, loss: 0.001289769890718162
step: 190, loss: 0.001302143675275147
step: 200, loss: 0.001138481660746038
step: 210, loss: 0.03973633050918579
step: 220, loss: 0.001974601997062564
step: 230, loss: 0.011341775767505169
/usr/local/lib/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=0.9910313901345291, f1=0.9831649831649831,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008124565938487649
step: 10, loss: 0.0008282199269160628
step: 20, loss: 0.0005554654053412378
step: 30, loss: 0.0006265910342335701
step: 40, loss: 0.0023777491878718138
step: 50, loss: 0.0011197729036211967
step: 60, loss: 0.00027895651874132454
step: 70, loss: 0.004285851493477821
step: 80, loss: 0.02755080536007881
step: 90, loss: 0.0035488305147737265
step: 100, loss: 0.004462207667529583
step: 110, loss: 0.00874937791377306
step: 120, loss: 0.03278280422091484
step: 130, loss: 0.02137788012623787
step: 140, loss: 0.004332792945206165
step: 150, loss: 0.0011183794122189283
step: 160, loss: 0.0003065946511924267
step: 170, loss: 0.023561399430036545
step: 180, loss: 0.010511280037462711
step: 190, loss: 0.0022499088663607836
step: 200, loss: 0.00032531225588172674
step: 210, loss: 0.00028204097179695964
step: 220, loss: 0.0003876641276292503
step: 230, loss: 0.00025018383166752756
/usr/local/lib/python3.7/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=0.9910112359550561, f1=0.9831649831649831,

```

```

best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022691093909088522
step: 10, loss: 0.0003350038605276495
step: 20, loss: 0.0001221374695887789
step: 30, loss: 0.00024153650156222284
step: 40, loss: 0.0002409225853625685
step: 50, loss: 0.0001655111846048385
step: 60, loss: 0.0003899928415194154
step: 70, loss: 0.0003453473618719727
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 4096.0
step: 80, loss: 0.00018297325004823506
step: 90, loss: 0.013067939318716526
step: 100, loss: 0.0038246596232056618
step: 110, loss: 0.0005083096912130713
step: 120, loss: 0.06258345395326614
step: 130, loss: 0.04705137386918068
step: 140, loss: 0.00045667539234273136
step: 150, loss: 0.00016998359933495522
step: 160, loss: 0.01832685060799122
step: 170, loss: 0.14476074278354645
step: 180, loss: 0.0002355246979277581
step: 190, loss: 0.0007320675649680197
step: 200, loss: 0.002307020127773285
step: 210, loss: 0.0013157152570784092
step: 220, loss: 0.0003275425697211176
step: 230, loss: 0.0003748202871065587
/usr/local/lib/python3.7/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=0.9877641824249165, f1=0.9800443458980044,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00034759624395519495
step: 10, loss: 0.07051318883895874
step: 20, loss: 0.00043919601012021303
step: 30, loss: 0.14436475932598114
step: 40, loss: 0.0005092337960377336
step: 50, loss: 0.0014040993992239237
step: 60, loss: 0.011815528385341167
step: 70, loss: 0.00036226067459210753
step: 80, loss: 0.0002220306487288326
step: 90, loss: 0.0005421940004453063
step: 100, loss: 0.0005020714015699923

```

```

step: 110, loss: 0.03591856732964516
step: 120, loss: 0.0002970259520225227
step: 130, loss: 0.0009524171473458409
step: 140, loss: 0.0012676379410549998
step: 150, loss: 0.0012300550006330013
step: 160, loss: 0.0016713703516870737
step: 170, loss: 0.00218234583735466
step: 180, loss: 0.0004454896552488208
step: 190, loss: 0.02042548544704914
step: 200, loss: 0.0007701037102378905
step: 210, loss: 0.0004815694992430508
step: 220, loss: 0.0002096585521940142
step: 230, loss: 0.004983667284250259
/usr/local/lib/python3.7/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=0.9875424688561721, f1=0.9853107344632768,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028576595708727837
step: 10, loss: 0.00018045640899799764
step: 20, loss: 0.00035084690898656845
step: 30, loss: 0.00013334964751265943
step: 40, loss: 0.0006212613079696894
step: 50, loss: 9.634006710257381e-05
step: 60, loss: 0.00043476696009747684
step: 70, loss: 0.00015192401770036668
step: 80, loss: 9.604595834389329e-05
step: 90, loss: 0.018619108945131302
step: 100, loss: 0.003247440094128251
step: 110, loss: 0.0086066834628582
step: 120, loss: 0.00015447422629222274
step: 130, loss: 0.00044625732698477805
step: 140, loss: 8.665648056194186e-05
step: 150, loss: 0.20075176656246185
step: 160, loss: 0.00023641811276320368
step: 170, loss: 0.0017293314449489117
step: 180, loss: 0.0011442683171480894
step: 190, loss: 0.0007391473045572639
step: 200, loss: 0.0005182810709811747
step: 210, loss: 0.0007644960423931479
step: 220, loss: 0.030972469598054886
step: 230, loss: 0.023309003561735153
/usr/local/lib/python3.7/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=0.9853768278965129, f1=0.980963045912654,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019650381058454514
step: 10, loss: 0.03647568076848984
step: 20, loss: 0.00032262393506243825
step: 30, loss: 0.00042705779196694493
step: 40, loss: 0.001959527377039194
step: 50, loss: 0.011159741319715977
step: 60, loss: 0.0014559895498678088
step: 70, loss: 0.00020599440904334188
step: 80, loss: 0.0005352976731956005
step: 90, loss: 0.00046511556138284504
step: 100, loss: 0.0005949363112449646
step: 110, loss: 0.00032126507721841335
step: 120, loss: 0.00012308852456044406
step: 130, loss: 0.0009440790163353086
step: 140, loss: 0.002254435094073415
step: 150, loss: 0.00017097311501856893
step: 160, loss: 0.0001776200660970062
step: 170, loss: 0.0001450442214263603
step: 180, loss: 0.0001668591285124421
step: 190, loss: 0.0011642160825431347
step: 200, loss: 0.2366732656955719
step: 210, loss: 0.0003036006528418511
step: 220, loss: 0.0003907516074832529
step: 230, loss: 0.01639203168451786
/usr/local/lib/python3.7/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=0.9910313901345291, f1=0.9810479375696767,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022937357425689697
step: 10, loss: 0.00037424947367981076
step: 20, loss: 0.001219558296725154
step: 30, loss: 0.0009477704297751188
step: 40, loss: 0.0001697889674687758
step: 50, loss: 0.00018755460041575134
step: 60, loss: 0.0004254784435033798
step: 70, loss: 0.0004856433079112321
step: 80, loss: 0.040431778877973557
step: 90, loss: 0.0003951993421651423
step: 100, loss: 0.00033604673808440566

```

```

step: 110, loss: 0.0019042453495785594
step: 120, loss: 0.060568686574697495
step: 130, loss: 0.00013682931603398174
step: 140, loss: 0.002750959014520049
step: 150, loss: 8.381364023080096e-05
step: 160, loss: 0.00025412969989702106
step: 170, loss: 0.0001067521734512411
step: 180, loss: 0.00015127891674637794
step: 190, loss: 7.707674376433715e-05
step: 200, loss: 0.0003692397731356323
step: 210, loss: 0.00013027008390054107
step: 220, loss: 0.0363159254193306
step: 230, loss: 0.006606850307434797
/usr/local/lib/python3.7/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=0.9898762654668166, f1=0.9831649831649831,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021709925204049796
step: 10, loss: 0.0001065274773281999
step: 20, loss: 0.0001503185776527971
step: 30, loss: 0.0001640495320316404
step: 40, loss: 5.901866825297475e-05
step: 50, loss: 0.00013875242439098656
step: 60, loss: 0.14268693327903748
step: 70, loss: 0.0005123773007653654
step: 80, loss: 0.006023809313774109
step: 90, loss: 0.00014262204058468342
step: 100, loss: 0.00033493523369543254
step: 110, loss: 0.00019268567848484963
step: 120, loss: 0.028760220855474472
step: 130, loss: 0.0007059655617922544
step: 140, loss: 0.0022025848738849163
step: 150, loss: 0.00026699810405261815
step: 160, loss: 0.00019879774481523782
step: 170, loss: 9.23827028600499e-05
step: 180, loss: 9.1530368081294e-05
step: 190, loss: 0.00013481917267199606
step: 200, loss: 7.84121875767596e-05
step: 210, loss: 9.662242518970743e-05
step: 220, loss: 0.0029845242388546467
step: 230, loss: 0.0001082095259334892
/usr/local/lib/python3.7/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=0.990990990990991, f1=0.9886877828054299,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.240709965117276e-05
step: 10, loss: 0.00010836481669684872
step: 20, loss: 4.455426096683368e-05
step: 30, loss: 7.984817057149485e-05
step: 40, loss: 0.003201324725523591
step: 50, loss: 0.0013873637653887272
step: 60, loss: 8.414722105953842e-05
step: 70, loss: 0.00016452591808047146
step: 80, loss: 0.00023077803780324757
step: 90, loss: 0.0002595882979221642
step: 100, loss: 0.00010230829502688721
step: 110, loss: 0.00021593106794171035
step: 120, loss: 0.00010147121065529063
step: 130, loss: 6.369647599058226e-05
step: 140, loss: 7.835988071747124e-05
step: 150, loss: 9.236687765223905e-05
step: 160, loss: 0.0006768340826965868
step: 170, loss: 0.005210087168961763
step: 180, loss: 0.00011863607505802065
step: 190, loss: 0.0001547669671708718
step: 200, loss: 0.00012733670882880688
step: 210, loss: 9.177317406283692e-05
step: 220, loss: 0.0002531585341785103
step: 230, loss: 0.01904352195560932
/usr/local/lib/python3.7/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=0.988814317673378, f1=0.984304932735426,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002961520804092288
step: 10, loss: 0.0009266362176276743
step: 20, loss: 7.573803304694593e-05
step: 30, loss: 0.0002168879727832973
step: 40, loss: 5.128217162564397e-05
step: 50, loss: 5.373269959818572e-05
step: 60, loss: 0.004751289263367653
step: 70, loss: 0.001043382566422224
step: 80, loss: 0.0009326220606453717
step: 90, loss: 7.354973786277696e-05
step: 100, loss: 6.0349895647959784e-05

```

```

step: 110, loss: 9.238816710421816e-05
step: 120, loss: 0.00012847574544139206
step: 130, loss: 0.0011702906340360641
step: 140, loss: 9.55900686676614e-05
step: 150, loss: 8.440502278972417e-05
step: 160, loss: 0.030371664091944695
step: 170, loss: 0.0002649631933309138
step: 180, loss: 0.00013702904107049108
step: 190, loss: 0.00023153664369601756
step: 200, loss: 0.0020059035159647465
step: 210, loss: 7.698180706938729e-05
step: 220, loss: 0.015573102980852127
step: 230, loss: 9.155426232609898e-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=0.9888641425389755, f1=0.9811738648947952,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003127592499367893
step: 10, loss: 0.00017451398889534175
step: 20, loss: 0.0001312654494540766
step: 30, loss: 0.010120381601154804
step: 40, loss: 0.00019117645570077002
step: 50, loss: 0.00011417699715821072
step: 60, loss: 5.8100798923987895e-05
step: 70, loss: 9.483774920227006e-05
step: 80, loss: 8.614992839284241e-05
step: 90, loss: 5.0054830353474244e-05
step: 100, loss: 0.00014104100409895182
step: 110, loss: 0.00016372438403777778
step: 120, loss: 8.718243770999834e-05
step: 130, loss: 0.0001044573073158972
step: 140, loss: 0.00010108685091836378
step: 150, loss: 0.012382508255541325
step: 160, loss: 5.657708607031964e-05
step: 170, loss: 0.00010652518540155143
step: 180, loss: 0.00018823746358975768
step: 190, loss: 5.246249202173203e-05
step: 200, loss: 7.302752055693418e-05
step: 210, loss: 0.00012652510486077517
step: 220, loss: 4.058798003825359e-05
step: 230, loss: 5.746654642280191e-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=0.9899216125419933, f1=0.9810901001112348,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.626866474514827e-05
step: 10, loss: 3.391616337466985e-05
step: 20, loss: 0.007503698114305735
step: 30, loss: 0.00014994748926255852
step: 40, loss: 4.279273343854584e-05
step: 50, loss: 0.00014319946058094501
step: 60, loss: 3.867344275931828e-05
step: 70, loss: 0.00011173773236805573
step: 80, loss: 0.00013242695422377437
step: 90, loss: 9.523804328637198e-05
step: 100, loss: 0.0008002538816072047
step: 110, loss: 0.00012215042079333216
step: 120, loss: 6.070742165320553e-05
step: 130, loss: 0.00014637017739005387
step: 140, loss: 9.183229121845216e-05
step: 150, loss: 8.838666690280661e-05
step: 160, loss: 9.455679537495598e-05
step: 170, loss: 8.687139779794961e-05
step: 180, loss: 0.0001141430766438134
step: 190, loss: 0.0007978577050380409
step: 200, loss: 6.144217331893742e-05
step: 210, loss: 0.017680982127785683
step: 220, loss: 8.662370237288997e-05
step: 230, loss: 3.194311648258008e-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=0.9898989898989898, f1=0.9832026875699889,
best_f1=0.9831649831649831
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.829966438701376e-05
step: 10, loss: 0.00015796021034475416
step: 20, loss: 9.178501204587519e-05
step: 30, loss: 9.37709555728361e-05
step: 40, loss: 9.092203254112974e-05
step: 50, loss: 0.0001244987506652251
step: 60, loss: 5.34706050530076e-05
step: 70, loss: 0.00012484921899158508
step: 80, loss: 0.02446937933564186
step: 90, loss: 4.34188186773099e-05
step: 100, loss: 9.881467849481851e-05

```

```

step: 110, loss: 7.504859968321398e-05
step: 120, loss: 9.733568731462583e-05
step: 130, loss: 9.656674228608608e-05
step: 140, loss: 8.677048026584089e-05
step: 150, loss: 0.00015865618479438126
step: 160, loss: 0.00015710710431449115
step: 170, loss: 5.4853862820891663e-05
step: 180, loss: 7.183389243436977e-05
step: 190, loss: 0.02761305309832096
step: 200, loss: 7.357278082054108e-05
step: 210, loss: 0.004477929323911667
step: 220, loss: 8.426512067671865e-05
step: 230, loss: 0.00013021349150221795
/usr/local/lib/python3.7/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=0.9898989898989898, f1=0.9854096520763187,
best_f1=0.9831649831649831
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, 235.61it/s]
load_f1 = 0.9887892376681614
real_f1 = 0.9876819708846584
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, 254.42it/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 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: 0.7917060852050781
step: 10, loss: 0.40508154034614563
step: 20, loss: 0.4607219994068146
step: 30, loss: 0.38227131962776184
step: 40, loss: 0.3142639100551605
step: 50, loss: 0.13220563530921936
step: 60, loss: 0.15515606105327606
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.13723161816596985
step: 80, loss: 0.043115679174661636
step: 90, loss: 0.06344334036111832
step: 100, loss: 0.4400646686553955
step: 110, loss: 0.07321646809577942
step: 120, loss: 0.016666721552610397
step: 130, loss: 0.03165720775723457
step: 140, loss: 0.20548216998577118

```



```

step: 150, loss: 0.027271103113889694
step: 160, loss: 0.1406291127204895
step: 170, loss: 0.16433708369731903
step: 180, loss: 0.10643112659454346
step: 190, loss: 0.03076913021504879
step: 200, loss: 0.10195476561784744
step: 210, loss: 0.07403965294361115
step: 220, loss: 0.03918502479791641
step: 230, loss: 0.08855350315570831
step: 240, loss: 0.12066002935171127
step: 250, loss: 0.031211109831929207
step: 260, loss: 0.02609539031982422
step: 270, loss: 0.005620163399726152
step: 280, loss: 0.07006724923849106
step: 290, loss: 0.038466908037662506
step: 300, loss: 0.17863236367702484
step: 310, loss: 0.05158169940114021
step: 320, loss: 0.06659262627363205
step: 330, loss: 0.06373032927513123
step: 340, loss: 0.11594223976135254
step: 350, loss: 0.09864078462123871
step: 360, loss: 0.057766567915678024
step: 370, loss: 0.11915435642004013
step: 380, loss: 0.12240128964185715
step: 390, loss: 0.015784043818712234
step: 400, loss: 0.012478209100663662
step: 410, loss: 0.018639659509062767
step: 420, loss: 0.010053119622170925
step: 430, loss: 0.08913306891918182
step: 440, loss: 0.05038256570696831
step: 450, loss: 0.013666239567101002
step: 460, loss: 0.1939251571893692
step: 470, loss: 0.1606636792421341
step: 480, loss: 0.3279978036880493
step: 490, loss: 0.02224799059331417
step: 500, loss: 0.006122818682342768
step: 510, loss: 0.03599684685468674
step: 520, loss: 0.07519736886024475
step: 530, loss: 0.02746722847223282
/usr/local/lib/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=0.9417206290471785, f1=0.9415016121602948,
best_f1=0.9415016121602948

```

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

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

```
step: 0, loss: 0.11392762511968613  
step: 10, loss: 0.14219483733177185  
step: 20, loss: 0.09833668917417526  
step: 30, loss: 0.03227008134126663  
step: 40, loss: 0.0021556676365435123  
step: 50, loss: 0.1572515219449997  
step: 60, loss: 0.1041257232427597  
step: 70, loss: 0.1527455896139145  
step: 80, loss: 0.01755894534289837  
step: 90, loss: 0.0023276687134057283  
step: 100, loss: 0.31974485516548157  
step: 110, loss: 0.023379480466246605  
step: 120, loss: 0.0351216197013855  
step: 130, loss: 0.031293176114559174  
step: 140, loss: 0.029426952823996544  
step: 150, loss: 0.09234044700860977  
step: 160, loss: 0.01932508498430252  
step: 170, loss: 0.09224411845207214  
step: 180, loss: 0.0016458292957395315  
step: 190, loss: 0.02201203443109989  
step: 200, loss: 0.0071921395137906075  
step: 210, loss: 0.010624062269926071  
step: 220, loss: 0.12245868146419525  
step: 230, loss: 0.025911984965205193  
step: 240, loss: 0.10382992774248123  
step: 250, loss: 0.008950320072472095  
step: 260, loss: 0.009586342610418797  
step: 270, loss: 0.13206173479557037  
step: 280, loss: 0.20542669296264648  
step: 290, loss: 0.08267650753259659  
step: 300, loss: 0.02594304084777832  
step: 310, loss: 0.06146274134516716  
step: 320, loss: 0.0847846195101738  
step: 330, loss: 0.012185025960206985  
step: 340, loss: 0.0095968097448349  
step: 350, loss: 0.03211703523993492  
step: 360, loss: 0.026975486427545547  
step: 370, loss: 0.0049108583480119705  
step: 380, loss: 0.056404516100883484  
step: 390, loss: 0.029531331732869148  
step: 400, loss: 0.027104854583740234  
step: 410, loss: 0.0006794119835831225  
step: 420, loss: 0.03464428707957268  
step: 430, loss: 0.0313827320933342  
step: 440, loss: 0.013818100094795227
```

```

step: 450, loss: 0.012926304712891579
step: 460, loss: 0.17516285181045532
step: 470, loss: 0.06996802240610123
step: 480, loss: 0.16691160202026367
step: 490, loss: 0.02764075994491577
step: 500, loss: 0.025555983185768127
step: 510, loss: 0.055825572460889816
step: 520, loss: 0.11112423241138458
step: 530, loss: 0.15787766873836517
/usr/local/lib/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=0.947906976744186, f1=0.9444444444444445,
best_f1=0.9444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008834215812385082
step: 10, loss: 0.03126494213938713
step: 20, loss: 0.17900817096233368
step: 30, loss: 0.2274656891822815
step: 40, loss: 0.0028747792821377516
step: 50, loss: 0.0046667298302054405
step: 60, loss: 0.0007323814206756651
step: 70, loss: 0.024803105741739273
step: 80, loss: 0.00037853041430935264
step: 90, loss: 0.07960769534111023
step: 100, loss: 0.015444563701748848
step: 110, loss: 0.014703563414514065
step: 120, loss: 0.005206857807934284
step: 130, loss: 0.026006150990724564
step: 140, loss: 0.03158954903483391
step: 150, loss: 0.0011004413245245814
step: 160, loss: 0.11246111243963242
step: 170, loss: 0.009254644624888897
step: 180, loss: 0.016361258924007416
step: 190, loss: 0.015652989968657494
step: 200, loss: 0.0006929713999852538
step: 210, loss: 0.009792591445147991
step: 220, loss: 0.030130336061120033
step: 230, loss: 0.013022823259234428
step: 240, loss: 0.01756352372467518
step: 250, loss: 0.00506154540926218
step: 260, loss: 0.001018025097437203

```

```

step: 270, loss: 0.0006704335100948811
step: 280, loss: 0.0016296359244734049
step: 290, loss: 0.039070919156074524
step: 300, loss: 0.07065626233816147
step: 310, loss: 0.046926241368055344
step: 320, loss: 0.14424702525138855
step: 330, loss: 0.0008415945339947939
step: 340, loss: 0.001277644420042634
step: 350, loss: 0.004115257877856493
step: 360, loss: 0.008570311591029167
step: 370, loss: 0.00948119256645441
step: 380, loss: 0.020518656820058823
step: 390, loss: 0.00427717249840498
step: 400, loss: 0.015577261336147785
step: 410, loss: 0.004416998941451311
step: 420, loss: 0.028316069394350052
step: 430, loss: 0.042597539722919464
step: 440, loss: 0.02736613154411316
step: 450, loss: 0.1419396847486496
step: 460, loss: 0.04146687686443329
step: 470, loss: 0.0027293965686112642
step: 480, loss: 0.002712859306484461
step: 490, loss: 0.01766037754714489
step: 500, loss: 0.04197102412581444
step: 510, loss: 0.0028447953518480062
step: 520, loss: 0.023001769557595253
step: 530, loss: 0.031208977103233337
/usr/local/lib/python3.7/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=0.9438629876308279, f1=0.9340344168260039,
best_f1=0.9444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008344080299139023
step: 10, loss: 0.023088760673999786
step: 20, loss: 0.09152988344430923
step: 30, loss: 0.0008616981795057654
step: 40, loss: 0.0005964056472294033
step: 50, loss: 0.006251834332942963
step: 60, loss: 0.000592369120568037
step: 70, loss: 0.006134715396910906
step: 80, loss: 0.004122037440538406
step: 90, loss: 0.029007794335484505
step: 100, loss: 0.05736047774553299
step: 110, loss: 0.012275965884327888
step: 120, loss: 0.06876170635223389

```

```

step: 130, loss: 0.04724156856536865
step: 140, loss: 0.05056184157729149
step: 150, loss: 0.0003225971304345876
step: 160, loss: 0.0015286022098734975
step: 170, loss: 0.0006661158986389637
step: 180, loss: 0.0009163320646621287
step: 190, loss: 0.05153818801045418
step: 200, loss: 0.0008458400843665004
step: 210, loss: 0.07970079034566879
step: 220, loss: 0.0012763338163495064
step: 230, loss: 0.19018013775348663
step: 240, loss: 0.006688016466796398
step: 250, loss: 0.0028725264128297567
step: 260, loss: 0.07950365543365479
step: 270, loss: 0.012467589229345322
step: 280, loss: 0.0017976443050429225
step: 290, loss: 0.03667280077934265
step: 300, loss: 0.0011326652020215988
step: 310, loss: 0.007438194938004017
step: 320, loss: 0.02952837385237217
step: 330, loss: 0.009638749063014984
step: 340, loss: 0.011323729529976845
step: 350, loss: 0.0002804627292789519
step: 360, loss: 0.0006675353506579995
step: 370, loss: 0.033283691853284836
step: 380, loss: 0.001577621791511774
step: 390, loss: 0.0229478906840086
step: 400, loss: 0.008431538008153439
step: 410, loss: 0.0031626771669834852
step: 420, loss: 0.009629834443330765
step: 430, loss: 0.007795749232172966
step: 440, loss: 0.026263907551765442
step: 450, loss: 0.006772888824343681
step: 460, loss: 0.003222750499844551
step: 470, loss: 0.002884737215936184
step: 480, loss: 0.001268703956156969
step: 490, loss: 0.005786763969808817
step: 500, loss: 0.0024445594754070044
step: 510, loss: 0.0779910609126091
step: 520, loss: 0.006208263803273439
step: 530, loss: 0.038180384784936905
/usr/local/lib/python3.7/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=0.943502824858757, f1=0.9354078264969354,
best_f1=0.9444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003067692741751671
step: 10, loss: 0.001709119649603963
step: 20, loss: 0.0010006459197029471
step: 30, loss: 0.00117088807746768
step: 40, loss: 0.03573455289006233
step: 50, loss: 0.008219449780881405
step: 60, loss: 0.0014861406525596976
step: 70, loss: 0.0002889807801693678
step: 80, loss: 0.0015864698216319084
step: 90, loss: 0.000886390043888241
step: 100, loss: 0.0005504531436599791
step: 110, loss: 0.007040537428110838
step: 120, loss: 0.0016673613572493196
step: 130, loss: 0.00042716722236946225
step: 140, loss: 0.00040082811028696597
step: 150, loss: 0.00018837174866348505
step: 160, loss: 0.03454690799117088
step: 170, loss: 0.01119379885494709
step: 180, loss: 0.050422150641679764
step: 190, loss: 0.00022919218463357538
step: 200, loss: 0.006063627079129219
step: 210, loss: 0.0037779139820486307
step: 220, loss: 0.00016623259580228478
step: 230, loss: 0.001545874634757638
step: 240, loss: 0.0002428734878776595
step: 250, loss: 9.409419726580381e-05
step: 260, loss: 0.00036794375046156347
step: 270, loss: 0.027200160548090935
step: 280, loss: 0.07679606229066849
step: 290, loss: 0.0002496226516086608
step: 300, loss: 0.0032054483890533447
step: 310, loss: 0.001564488047733903
step: 320, loss: 0.004058075603097677
step: 330, loss: 0.0016328425845131278
step: 340, loss: 0.0007298963028006256
step: 350, loss: 0.0002630694943945855
step: 360, loss: 0.0024157478474080563
step: 370, loss: 0.0013925540260970592
step: 380, loss: 0.0340411551296711
step: 390, loss: 0.046826086938381195
step: 400, loss: 0.15348565578460693
step: 410, loss: 0.00019197216897737235
step: 420, loss: 0.0027620296459645033
step: 430, loss: 0.0034926999360322952
step: 440, loss: 0.0015175711596384645
step: 450, loss: 0.00023776781745254993
step: 460, loss: 0.0072942934930324554
```

```

step: 470, loss: 0.0017323129577562213
step: 480, loss: 0.0024682399816811085
step: 490, loss: 0.002422113437205553
step: 500, loss: 0.005024867597967386
step: 510, loss: 0.007109650410711765
step: 520, loss: 0.00022900404292158782
step: 530, loss: 0.03858806937932968
/usr/local/lib/python3.7/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=0.9470260223048327, f1=0.937995337995338,
best_f1=0.9444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008250407874584198
step: 10, loss: 0.002089431742206216
step: 20, loss: 0.0002810711448546499
step: 30, loss: 0.06340329349040985
step: 40, loss: 0.000219455745536834
step: 50, loss: 0.0046746088191866875
step: 60, loss: 0.00024063307500910014
step: 70, loss: 0.00160779629368335
step: 80, loss: 0.0009390506893396378
step: 90, loss: 0.0006175939925014973
step: 100, loss: 0.01145476009696722
step: 110, loss: 0.008539136499166489
step: 120, loss: 0.00021433828806038946
step: 130, loss: 0.0001264798193005845
step: 140, loss: 0.0007632820052094758
step: 150, loss: 0.00033281900687143207
step: 160, loss: 8.276657899841666e-05
step: 170, loss: 6.579836190212518e-05
step: 180, loss: 8.456045179627836e-05
step: 190, loss: 9.08591173356399e-05
step: 200, loss: 0.00011755530431400985
step: 210, loss: 0.002103084698319435
step: 220, loss: 0.00028037320589646697
step: 230, loss: 6.58650096738711e-05
step: 240, loss: 0.00501724099740386
step: 250, loss: 0.00032172707142308354
step: 260, loss: 0.0007741536828689277
step: 270, loss: 0.0006636274047195911
step: 280, loss: 0.0020238158758729696
step: 290, loss: 0.0028545414097607136
step: 300, loss: 0.0391785092651844
step: 310, loss: 0.0011550611816346645
step: 320, loss: 0.010114234872162342

```

```

step: 330, loss: 0.002836567349731922
step: 340, loss: 0.0012560460017994046
step: 350, loss: 0.0531780980527401
step: 360, loss: 0.019812971353530884
step: 370, loss: 0.0003741858236026019
step: 380, loss: 0.0031016927678138018
step: 390, loss: 0.0016972740413621068
step: 400, loss: 0.00012666746624745429
step: 410, loss: 5.758035695180297e-05
step: 420, loss: 0.007105071097612381
step: 430, loss: 0.00019560795044526458
step: 440, loss: 0.002142152516171336
step: 450, loss: 0.13796934485435486
step: 460, loss: 0.009137192741036415
step: 470, loss: 0.029655149206519127
step: 480, loss: 0.03531015291810036
step: 490, loss: 0.00020087993470951915
step: 500, loss: 0.0015618063043802977
step: 510, loss: 0.0002595371042843908
step: 520, loss: 0.0016121735097840428
step: 530, loss: 0.0022684053983539343
/usr/local/lib/python3.7/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=0.9473684210526315, f1=0.9389493610979649,
best_f1=0.9444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002386263106018305
step: 10, loss: 0.003805827582255006
step: 20, loss: 0.0033924931194633245
step: 30, loss: 0.005873099900782108
step: 40, loss: 8.14262093626894e-05
step: 50, loss: 0.0002684720966499299
step: 60, loss: 0.000815300561953336
step: 70, loss: 0.0005718620377592742
step: 80, loss: 9.228336421074346e-05
step: 90, loss: 6.959238817216828e-05
step: 100, loss: 0.0003300889802630991
step: 110, loss: 0.04761086031794548
step: 120, loss: 5.996919207973406e-05
step: 130, loss: 0.00025629825540818274
step: 140, loss: 0.0014506556326523423
step: 150, loss: 0.0005431463359855115
step: 160, loss: 0.0006440442521125078
step: 170, loss: 0.00031598570058122277
step: 180, loss: 0.0025389764923602343

```



```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.0001181808693218045
step: 200, loss: 0.00014196569100022316
step: 210, loss: 0.00017174892127513885
step: 220, loss: 0.0002015705540543422
step: 230, loss: 0.00026968595921061933
step: 240, loss: 0.0018060116562992334
step: 250, loss: 0.0034138213377445936
step: 260, loss: 0.0002627514477353543
step: 270, loss: 8.203856850741431e-05
step: 280, loss: 0.005398145876824856
step: 290, loss: 0.005192680284380913
step: 300, loss: 0.03749242424964905
step: 310, loss: 0.0003732965560629964
step: 320, loss: 0.01803576946258545
step: 330, loss: 0.0006512449472211301
step: 340, loss: 0.0008503368007950485
step: 350, loss: 0.00023013827740214765
step: 360, loss: 0.003601194592192769
step: 370, loss: 0.005923599470406771
step: 380, loss: 0.0018497378332540393
step: 390, loss: 0.0004008478135801852
step: 400, loss: 6.0784608649555594e-05
step: 410, loss: 0.0003632156876847148
step: 420, loss: 0.0004713111266028136
step: 430, loss: 6.504235352622345e-05
step: 440, loss: 0.00011169204663019627
step: 450, loss: 0.00021304113033693284
step: 460, loss: 0.0015331762842833996
step: 470, loss: 0.0020395396277308464
step: 480, loss: 0.0005279593751765788
step: 490, loss: 0.004319266881793737
step: 500, loss: 6.255839980440214e-05
step: 510, loss: 0.0004423802311066538
step: 520, loss: 0.0009628443513065577
step: 530, loss: 6.67451968183741e-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 7: dev_f1=0.9489939167056621, f1=0.9370892018779343,
best_f1=0.9370892018779343
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.0008443731931038201
step: 10, loss: 0.0031876908615231514
step: 20, loss: 0.06555299460887909
step: 30, loss: 0.0017708708764985204
step: 40, loss: 0.0002507873286958784
step: 50, loss: 0.00017117599782068282
step: 60, loss: 0.0021923824679106474
step: 70, loss: 0.0010650536278262734
step: 80, loss: 0.0005689229001291096
step: 90, loss: 5.71187847526744e-05
step: 100, loss: 0.0005837624194100499
step: 110, loss: 0.000246119627263397
step: 120, loss: 0.003942409995943308
step: 130, loss: 0.0002479791874065995
step: 140, loss: 3.214053504052572e-05
step: 150, loss: 0.0005327578983269632
step: 160, loss: 0.00014562663272954524
step: 170, loss: 0.001824419479817152
step: 180, loss: 0.0004098792851436883
step: 190, loss: 0.0003617892798501998
step: 200, loss: 0.0002702216152101755
step: 210, loss: 0.00014880907838232815
step: 220, loss: 0.002001018961891532
step: 230, loss: 0.00012212446017656475
step: 240, loss: 8.78529972396791e-05
step: 250, loss: 0.0068415808491408825
step: 260, loss: 0.0032005265820771456
step: 270, loss: 4.4060099753551185e-05
step: 280, loss: 0.00012732126924674958
step: 290, loss: 0.00048674814752303064
step: 300, loss: 4.186669320915826e-05
step: 310, loss: 0.06616207957267761
step: 320, loss: 0.00018564777565188706
step: 330, loss: 0.016958720982074738
step: 340, loss: 0.002082924358546734
step: 350, loss: 0.03860095143318176
step: 360, loss: 0.0031151913572102785
step: 370, loss: 0.0016659500543028116
step: 380, loss: 0.00558607978746295
step: 390, loss: 0.0008534969529137015
step: 400, loss: 0.0023461191449314356
step: 410, loss: 0.0005478714010678232
step: 420, loss: 5.990083081997e-05
step: 430, loss: 9.537757432553917e-05
step: 440, loss: 0.00011202166933799163
step: 450, loss: 0.0008397463243454695
step: 460, loss: 0.002926729153841734
step: 470, loss: 0.00010812435357365757

```

step: 480, loss: 0.00046291612670756876
step: 490, loss: 7.88276011007838e-05
step: 500, loss: 0.009539974853396416
step: 510, loss: 4.5328812120715156e-05
step: 520, loss: 7.330704829655588e-05
step: 530, loss: 4.2383966501802206e-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 8: dev_f1=0.9493258949325895, f1=0.9459084604715674,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00024203784414567053
step: 10, loss: 0.18659980595111847
step: 20, loss: 0.00013222455163486302
step: 30, loss: 0.002099004341289401
step: 40, loss: 0.0004384623025543988
step: 50, loss: 0.0002143009623978287
step: 60, loss: 0.0008898327359929681
step: 70, loss: 0.10787712037563324
step: 80, loss: 8.460104436380789e-05
step: 90, loss: 0.001470965682528913
step: 100, loss: 0.0005633849650621414
step: 110, loss: 0.002306309062987566
step: 120, loss: 6.296714855125174e-05
step: 130, loss: 0.00015591712144669145
step: 140, loss: 0.00040480561438016593
step: 150, loss: 0.00010547334386501461
step: 160, loss: 0.0005720985354855657
step: 170, loss: 5.274188515613787e-05
step: 180, loss: 0.00018476184050086886
step: 190, loss: 0.0003080348833464086
step: 200, loss: 6.135306466603652e-05
step: 210, loss: 0.00020744405628647655
step: 220, loss: 0.00010520947398617864
step: 230, loss: 3.113119601039216e-05
step: 240, loss: 6.923818727955222e-05
step: 250, loss: 0.004120754078030586
step: 260, loss: 0.00044136549695394933
step: 270, loss: 0.00022817065473645926
step: 280, loss: 4.442560748429969e-05
step: 290, loss: 3.944916898035444e-05

```

```

step: 300, loss: 0.00038651638897135854
step: 310, loss: 3.0159015295794234e-05
step: 320, loss: 0.008829430676996708
step: 330, loss: 3.621232826844789e-05
step: 340, loss: 4.2019968532258645e-05
step: 350, loss: 6.548514647874981e-05
step: 360, loss: 0.017323654145002365
step: 370, loss: 9.418428089702502e-05
step: 380, loss: 3.4029075322905555e-05
step: 390, loss: 0.019791940227150917
step: 400, loss: 0.06275336444377899
step: 410, loss: 0.0006352383643388748
step: 420, loss: 0.0007868175162002444
step: 430, loss: 4.042783984914422e-05
step: 440, loss: 0.0001583535922691226
step: 450, loss: 0.011154409497976303
step: 460, loss: 0.00038153628702275455
step: 470, loss: 0.0005875904462300241
step: 480, loss: 0.0005418814835138619
step: 490, loss: 6.098561789258383e-05
step: 500, loss: 0.007809820119291544
step: 510, loss: 0.010509971529245377
step: 520, loss: 0.00119981134776026
step: 530, loss: 0.0008160935831256211
/usr/local/lib/python3.7/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=0.9481065918653577, f1=0.9447565543071161,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006812494248151779
step: 10, loss: 3.139472391922027e-05
step: 20, loss: 0.00010997508798027411
step: 30, loss: 5.681282709701918e-05
step: 40, loss: 0.00024060197756625712
step: 50, loss: 0.00016779477300588042
step: 60, loss: 0.0007692899671383202
step: 70, loss: 0.003192741656675935
step: 80, loss: 0.0008922192500904202
step: 90, loss: 5.343026350601576e-05
step: 100, loss: 6.114951975177974e-05
step: 110, loss: 0.0008714103605598211
step: 120, loss: 4.5054126530885696e-05
step: 130, loss: 4.4862914364784956e-05
step: 140, loss: 5.9042191423941404e-05
step: 150, loss: 4.3490326788742095e-05

```

```

step: 160, loss: 0.0008859772933647037
step: 170, loss: 0.00019536026229616255
step: 180, loss: 0.001612690626643598
step: 190, loss: 9.178027539746836e-05
step: 200, loss: 0.000236944601056166
step: 210, loss: 0.0025522520300000906
step: 220, loss: 0.000481412309454754
step: 230, loss: 4.3776395614258945e-05
step: 240, loss: 3.0639446777058765e-05
step: 250, loss: 0.00037659110967069864
step: 260, loss: 0.0005117857363075018
step: 270, loss: 0.012757541611790657
step: 280, loss: 0.0002470863691996783
step: 290, loss: 0.0005297334864735603
step: 300, loss: 0.00010978429054375738
step: 310, loss: 0.00014778749027755111
step: 320, loss: 0.00021817997912876308
step: 330, loss: 0.00011979834380326793
step: 340, loss: 0.00017089067841880023
step: 350, loss: 3.2733030820963904e-05
step: 360, loss: 0.00011071567132603377
step: 370, loss: 0.00043224202818237245
step: 380, loss: 0.00018747392459772527
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 390, loss: 3.19246064464096e-05
step: 400, loss: 4.039429040858522e-05
step: 410, loss: 0.0036650162655860186
step: 420, loss: 4.698928751167841e-05
step: 430, loss: 3.477016798569821e-05
step: 440, loss: 3.228964487789199e-05
step: 450, loss: 0.0012370033655315638
step: 460, loss: 0.00024863192811608315
step: 470, loss: 2.9518292649299838e-05
step: 480, loss: 7.112856837920845e-05
step: 490, loss: 0.036833133548498154
step: 500, loss: 0.00020734310965053737
step: 510, loss: 0.00012108348164474592
step: 520, loss: 0.00010624605056364089
step: 530, loss: 0.0005193756660446525
/usr/local/lib/python3.7/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=0.9474174034434621, f1=0.943466172381835,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.633832087274641e-05

```

step: 10, loss: 0.0010445197112858295
step: 20, loss: 2.497342757123988e-05
step: 30, loss: 0.002455186564475298
step: 40, loss: 6.947126530576497e-05
step: 50, loss: 4.235046071698889e-05
step: 60, loss: 8.775711467023939e-05
step: 70, loss: 9.080286690732464e-05
step: 80, loss: 7.565703708678484e-05
step: 90, loss: 7.712790102232248e-05
step: 100, loss: 5.800451981485821e-05
step: 110, loss: 4.8251644329866394e-05
step: 120, loss: 8.172771049430594e-05
step: 130, loss: 4.299863212509081e-05
step: 140, loss: 2.495421540515963e-05
step: 150, loss: 0.0001395284489262849
step: 160, loss: 4.660795093514025e-05
step: 170, loss: 9.161604975815862e-05
step: 180, loss: 9.534628043184057e-05
step: 190, loss: 0.00011486233415780589
step: 200, loss: 0.0013881274498999119
step: 210, loss: 4.304324829718098e-05
step: 220, loss: 3.442374509177171e-05
step: 230, loss: 3.279189331806265e-05
step: 240, loss: 7.074450695654377e-05
step: 250, loss: 2.6843212253879756e-05
step: 260, loss: 3.172319338773377e-05
step: 270, loss: 0.0003339464601594955
step: 280, loss: 8.71839583851397e-05
step: 290, loss: 0.022955799475312233
step: 300, loss: 0.0698935016989708
step: 310, loss: 5.447400326374918e-05
step: 320, loss: 0.02529594488441944
step: 330, loss: 0.0022946030367165804
step: 340, loss: 4.8530706408200786e-05
step: 350, loss: 0.00011670839739963412
step: 360, loss: 4.055053068441339e-05
step: 370, loss: 2.1758836737717502e-05
step: 380, loss: 3.801007551373914e-05
step: 390, loss: 0.020693164318799973
step: 400, loss: 3.0320761652546935e-05
step: 410, loss: 2.257470623590052e-05
step: 420, loss: 4.230946797179058e-05
step: 430, loss: 5.301505734678358e-05
step: 440, loss: 0.00012298607907723635
step: 450, loss: 2.9759878088952973e-05
step: 460, loss: 0.001123074209317565
step: 470, loss: 0.02131466567516327
step: 480, loss: 0.0003778446407523006

```

step: 490, loss: 0.0001234813971677795
step: 500, loss: 0.00019183213589712977
step: 510, loss: 2.6746740331873298e-05
step: 520, loss: 2.5971743525587954e-05
step: 530, loss: 3.137518069706857e-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=0.9469234382339127, f1=0.9477611940298507,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.744852074305527e-05
step: 10, loss: 4.5039367250865325e-05
step: 20, loss: 3.177011967636645e-05
step: 30, loss: 2.7033685910282657e-05
step: 40, loss: 0.00013033440336585045
step: 50, loss: 0.0008064596331678331
step: 60, loss: 0.00040698700468055904
step: 70, loss: 6.490943633252755e-05
step: 80, loss: 0.001471970696002245
step: 90, loss: 3.0315268304548226e-05
step: 100, loss: 4.9617628974374384e-05
step: 110, loss: 3.378505789441988e-05
step: 120, loss: 6.968530942685902e-05
step: 130, loss: 4.3965494114672765e-05
step: 140, loss: 3.0084309400990605e-05
step: 150, loss: 5.2544899517670274e-05
step: 160, loss: 7.436671876348555e-05
step: 170, loss: 0.00021379523968789726
step: 180, loss: 0.00025102103245444596
step: 190, loss: 0.0024980928283184767
step: 200, loss: 5.301878263708204e-05
step: 210, loss: 0.00011746265954570845
step: 220, loss: 2.5268020181101747e-05
step: 230, loss: 0.0012392211938276887
step: 240, loss: 0.0027630142867565155
step: 250, loss: 0.0001645463053137064
step: 260, loss: 0.0005900576943531632
step: 270, loss: 0.13382476568222046
step: 280, loss: 2.9644481401192024e-05
step: 290, loss: 0.00640897499397397
step: 300, loss: 2.9517974326154217e-05
step: 310, loss: 2.579306783445645e-05
step: 320, loss: 0.00012296726345084608
step: 330, loss: 6.270999438129365e-05
step: 340, loss: 0.006337752565741539

```

```

step: 350, loss: 0.017874693498015404
step: 360, loss: 0.0002757053298410028
step: 370, loss: 2.0931718609062955e-05
step: 380, loss: 3.0046905521885492e-05
step: 390, loss: 3.3506155887153e-05
step: 400, loss: 2.382991988270078e-05
step: 410, loss: 5.5201275245053694e-05
step: 420, loss: 0.0006010967190377414
step: 430, loss: 5.492740456247702e-05
step: 440, loss: 0.0017427445854991674
step: 450, loss: 0.0036515311803668737
step: 460, loss: 7.113238825695589e-05
step: 470, loss: 2.8045295039191842e-05
step: 480, loss: 0.00010750415822258219
step: 490, loss: 0.04468673840165138
step: 500, loss: 0.0006319903186522424
step: 510, loss: 3.43457140843384e-05
step: 520, loss: 0.017614537850022316
step: 530, loss: 0.00017269425734411925
/usr/local/lib/python3.7/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=0.9492619926199262, f1=0.9462068965517241,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.912666943506338e-05
step: 10, loss: 2.096496791637037e-05
step: 20, loss: 0.034360308200120926
step: 30, loss: 0.001909833517856896
step: 40, loss: 0.001817109645344317
step: 50, loss: 3.687591015477665e-05
step: 60, loss: 4.470210842555389e-05
step: 70, loss: 0.004147885832935572
step: 80, loss: 2.103994665958453e-05
step: 90, loss: 0.000952998991124332
step: 100, loss: 2.1006020688218996e-05
step: 110, loss: 1.736699960019905e-05
step: 120, loss: 3.098585148109123e-05
step: 130, loss: 0.000830001023132354
step: 140, loss: 8.319792686961591e-05
step: 150, loss: 3.829701745416969e-05
step: 160, loss: 2.471248444635421e-05
step: 170, loss: 6.594845035579056e-05
step: 180, loss: 2.6604606318869628e-05
step: 190, loss: 0.00013354270777199417
step: 200, loss: 0.002427069004625082

```



```

step: 210, loss: 0.0005637300782836974
step: 220, loss: 0.00042998031130991876
step: 230, loss: 0.0036485325545072556
step: 240, loss: 3.239012949052267e-05
step: 250, loss: 0.044060979038476944
step: 260, loss: 4.193168570054695e-05
step: 270, loss: 1.9657769371406175e-05
step: 280, loss: 1.3150125596439466e-05
step: 290, loss: 0.00012338953092694283
step: 300, loss: 3.525035208440386e-05
step: 310, loss: 1.7836367987911217e-05
step: 320, loss: 0.00012727515422739089
step: 330, loss: 7.959555659908801e-05
step: 340, loss: 3.61766797141172e-05
step: 350, loss: 0.000617838988546282
step: 360, loss: 2.007881812460255e-05
step: 370, loss: 4.9071895773522556e-05
step: 380, loss: 0.0002448191517032683
step: 390, loss: 0.00014983527944423258
step: 400, loss: 1.9650306057883427e-05
step: 410, loss: 0.0010320032015442848
step: 420, loss: 5.330010753823444e-05
step: 430, loss: 0.00035447755362838507
step: 440, loss: 0.0013505963142961264
step: 450, loss: 2.5420536985620856e-05
step: 460, loss: 2.5021450710482895e-05
step: 470, loss: 0.025396134704351425
step: 480, loss: 0.03884749859571457
step: 490, loss: 0.00028626355924643576
step: 500, loss: 4.3672134779626504e-05
step: 510, loss: 3.45424996339716e-05
step: 520, loss: 2.601992127893027e-05
step: 530, loss: 0.0006728356820531189
/usr/local/lib/python3.7/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=0.9455399061032863, f1=0.9443665264142123,
best_f1=0.9459084604715674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.526780554035213e-05
step: 10, loss: 4.756015187012963e-05
step: 20, loss: 2.2834439732832834e-05
step: 30, loss: 4.694039671448991e-05
step: 40, loss: 0.0002473889326211065
step: 50, loss: 2.9826956961187534e-05
step: 60, loss: 2.4988212317111902e-05

```

step: 70, loss: 2.0309713363531046e-05
step: 80, loss: 0.0020431512966752052
step: 90, loss: 1.3086788385408e-05
step: 100, loss: 3.459870276856236e-05
step: 110, loss: 1.999675077968277e-05
step: 120, loss: 3.435015969444066e-05
step: 130, loss: 0.0006595306331291795
step: 140, loss: 0.000348778412444517
step: 150, loss: 5.112207873025909e-05
step: 160, loss: 0.0017502270638942719
step: 170, loss: 2.752504042291548e-05
step: 180, loss: 0.0009037883137352765
step: 190, loss: 4.1908377170329913e-05
step: 200, loss: 0.00016127375420182943
step: 210, loss: 0.00014294468564912677
step: 220, loss: 1.721799526421819e-05
step: 230, loss: 0.002595706842839718
step: 240, loss: 2.3606387912877835e-05
step: 250, loss: 6.503018084913492e-05
step: 260, loss: 1.2643488844332751e-05
step: 270, loss: 0.005074895452708006
step: 280, loss: 4.000299304607324e-05
step: 290, loss: 0.002730065491050482
step: 300, loss: 2.0037632566527463e-05
step: 310, loss: 2.884870264097117e-05
step: 320, loss: 9.115917782764882e-05
step: 330, loss: 0.0002281791967106983
step: 340, loss: 3.413302329136059e-05
step: 350, loss: 6.606747774640098e-05
step: 360, loss: 0.0036588991060853004
step: 370, loss: 2.2958261979511008e-05
step: 380, loss: 0.00010158078657696024
step: 390, loss: 2.3446093109669164e-05
step: 400, loss: 3.791307972278446e-05
step: 410, loss: 2.04774896701565e-05
step: 420, loss: 2.2529575289809145e-05
step: 430, loss: 4.320883090258576e-05
step: 440, loss: 1.4606633158109616e-05
step: 450, loss: 2.881672298826743e-05
step: 460, loss: 0.000115201786684338
step: 470, loss: 1.7225256669917144e-05
step: 480, loss: 1.910656465042848e-05
step: 490, loss: 2.2216383513296023e-05
step: 500, loss: 2.5517174435663037e-05
step: 510, loss: 0.0013733139494434
step: 520, loss: 4.124944825889543e-05
step: 530, loss: 2.686968036869075e-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=0.949438202247191, f1=0.946236559139785,
best_f1=0.946236559139785
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 1.8793431081576273e-05
step: 10, loss: 0.0005659413873218
step: 20, loss: 2.7729460271075368e-05
step: 30, loss: 3.718500374816358e-05
step: 40, loss: 1.4379379535967018e-05
step: 50, loss: 0.000414088397519663
step: 60, loss: 1.5593739590258338e-05
step: 70, loss: 5.5708900617901236e-05
step: 80, loss: 0.00011229480151087046
step: 90, loss: 1.9750841602217406e-05
step: 100, loss: 2.3632135707885027e-05
step: 110, loss: 0.013089757412672043
step: 120, loss: 0.0021332509350031614
step: 130, loss: 2.8021862817695364e-05
step: 140, loss: 1.4651340279669967e-05
step: 150, loss: 0.0012928565265610814
step: 160, loss: 5.270808833301999e-05
step: 170, loss: 2.0373205188661814e-05
step: 180, loss: 2.5941446438082494e-05
step: 190, loss: 0.00015103512851055712
step: 200, loss: 2.2425654606195167e-05
step: 210, loss: 1.5042457562230993e-05
step: 220, loss: 1.2501901437644847e-05
step: 230, loss: 2.4932041924330406e-05
step: 240, loss: 2.4075992769212462e-05
step: 250, loss: 0.00027309736469760537
step: 260, loss: 3.993724749307148e-05
step: 270, loss: 8.333038567798212e-05
step: 280, loss: 0.000645998225081712
step: 290, loss: 0.002398098586127162
step: 300, loss: 0.001077475375495851
step: 310, loss: 2.105477688019164e-05
step: 320, loss: 3.9030128391459584e-05
step: 330, loss: 4.5623339246958494e-05
step: 340, loss: 0.00011792818986577913
step: 350, loss: 5.4295709560392424e-05
step: 360, loss: 1.9028417227673344e-05

```

```

step: 370, loss: 1.6867814338183962e-05
step: 380, loss: 1.5616180462529883e-05
step: 390, loss: 1.548206273582764e-05
step: 400, loss: 1.5683219316997565e-05
step: 410, loss: 0.00043919371091760695
step: 420, loss: 2.294356090715155e-05
step: 430, loss: 1.7709746316540986e-05
step: 440, loss: 0.0008060596883296967
step: 450, loss: 0.0011466998839750886
step: 460, loss: 1.5131903637666255e-05
step: 470, loss: 7.018764154054224e-05
step: 480, loss: 2.141955519618932e-05
step: 490, loss: 3.7928934034425765e-05
step: 500, loss: 0.00021172492415644228
step: 510, loss: 2.8183774702483788e-05
step: 520, loss: 6.684412073809654e-05
step: 530, loss: 3.01395557471551e-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=0.9491367242183854, f1=0.9456066945606694,
best_f1=0.946236559139785
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, 258.25it/s]
load_f1 = 0.9504015115729806
real_f1 = 0.947219604147031
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, 253.64it/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: 0.8519405722618103
```

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

```
step: 10, loss: 0.0637040063738823
```

```
step: 20, loss: 0.3835756480693817
```

```
step: 30, loss: 0.36930903792381287
```

```
step: 40, loss: 0.5014875531196594
```

```
step: 50, loss: 0.2908928096294403
```

```
step: 60, loss: 0.34474658966064453
```

```
step: 70, loss: 0.1798432171344757
```

```
step: 80, loss: 0.3195801079273224
```

```
step: 90, loss: 0.3546658158302307
```

```
step: 100, loss: 0.11144442856311798
```

```

step: 110, loss: 0.2537016272544861
step: 120, loss: 0.17220555245876312
step: 130, loss: 0.17327596247196198
step: 140, loss: 0.15758365392684937
step: 150, loss: 0.2568204700946808
step: 160, loss: 0.17786799371242523
step: 170, loss: 0.06351414322853088
step: 180, loss: 0.17483201622962952
step: 190, loss: 0.22692729532718658
step: 200, loss: 0.14225265383720398
step: 210, loss: 0.3351781368255615
/usr/local/lib/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=0.6850828729281767, f1=0.6988416988416989,
best_f1=0.6988416988416989
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07993412762880325
step: 10, loss: 0.05464371293783188
step: 20, loss: 0.16349968314170837
step: 30, loss: 0.07632127404212952
step: 40, loss: 0.07146593928337097
step: 50, loss: 0.1723749190568924
step: 60, loss: 0.06697577238082886
step: 70, loss: 0.13766610622406006
step: 80, loss: 0.20751523971557617
step: 90, loss: 0.04939035698771477
step: 100, loss: 0.1474643349647522
step: 110, loss: 0.058344483375549316
step: 120, loss: 0.15763135254383087
step: 130, loss: 0.1422690898180008
step: 140, loss: 0.11035457253456116
step: 150, loss: 0.12291014939546585
step: 160, loss: 0.10224610567092896
step: 170, loss: 0.15192581713199615
step: 180, loss: 0.16218367218971252
step: 190, loss: 0.08453048765659332
step: 200, loss: 0.12159313261508942
step: 210, loss: 0.05588991194963455
/usr/local/lib/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=0.7079303675048356, f1=0.7123287671232877,
best_f1=0.7123287671232877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12125562131404877
step: 10, loss: 0.1539355218410492
step: 20, loss: 0.225501149892807
step: 30, loss: 0.07125423848628998
step: 40, loss: 0.16305284202098846
step: 50, loss: 0.07462514936923981
step: 60, loss: 0.09083716571331024
step: 70, loss: 0.0850105881690979
step: 80, loss: 0.09471224248409271
step: 90, loss: 0.11739636212587357
step: 100, loss: 0.010795846581459045
step: 110, loss: 0.06270293891429901
step: 120, loss: 0.27198535203933716
step: 130, loss: 0.0954071432352066
step: 140, loss: 0.11497402936220169
step: 150, loss: 0.09818863868713379
step: 160, loss: 0.0515783317387104
step: 170, loss: 0.05764327943325043
step: 180, loss: 0.027774455025792122
step: 190, loss: 0.09667645394802094
step: 200, loss: 0.08057523518800735
step: 210, loss: 0.14868654310703278
/usr/local/lib/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=0.7166979362101313, f1=0.7448789571694598,
best_f1=0.7448789571694598
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.052482787519693375
step: 10, loss: 0.020364413037896156
step: 20, loss: 0.042396653443574905
step: 30, loss: 0.049129847437143326
step: 40, loss: 0.009018873795866966

```



```

step: 50, loss: 0.04466455802321434
step: 60, loss: 0.19965240359306335
step: 70, loss: 0.07579591870307922
step: 80, loss: 0.09929429739713669
step: 90, loss: 0.019827358424663544
step: 100, loss: 0.018635563552379608
step: 110, loss: 0.06503065675497055
step: 120, loss: 0.2128899097442627
step: 130, loss: 0.12296963483095169
step: 140, loss: 0.11147312819957733
step: 150, loss: 0.07672785967588425
step: 160, loss: 0.020576927810907364
step: 170, loss: 0.02364264614880085
step: 180, loss: 0.10858136415481567
step: 190, loss: 0.05289270356297493
step: 200, loss: 0.016748161986470222
step: 210, loss: 0.0134472930803895
/usr/local/lib/python3.7/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=0.7054263565891473, f1=0.7094339622641509,
best_f1=0.7448789571694598
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.035567425191402435
step: 10, loss: 0.08324578404426575
step: 20, loss: 0.02873845584690571
step: 30, loss: 0.04981745034456253
step: 40, loss: 0.046531032770872116
step: 50, loss: 0.08662915229797363
step: 60, loss: 0.01882070116698742
step: 70, loss: 0.04267221689224243
step: 80, loss: 0.010284154675900936
step: 90, loss: 0.05395825579762459
step: 100, loss: 0.03916456177830696
step: 110, loss: 0.010998462326824665
step: 120, loss: 0.004161291755735874
step: 130, loss: 0.01953018456697464
step: 140, loss: 0.03165043890476227
step: 150, loss: 0.08667689561843872
step: 160, loss: 0.08017858862876892
step: 170, loss: 0.02295728772878647
step: 180, loss: 0.062058690935373306
step: 190, loss: 0.0419219508767128
step: 200, loss: 0.09529457986354828
step: 210, loss: 0.012790383771061897
/usr/local/lib/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=0.7290448343079923, f1=0.7347740667976425,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11322194337844849
step: 10, loss: 0.05047665536403656
step: 20, loss: 0.0066973427310585976
step: 30, loss: 0.07292618602514267
step: 40, loss: 0.003634373191744089
step: 50, loss: 0.021247658878564835
step: 60, loss: 0.17116567492485046
step: 70, loss: 0.0019300789572298527
step: 80, loss: 0.1112377718091011
step: 90, loss: 0.01931372657418251
step: 100, loss: 0.021032609045505524
step: 110, loss: 0.08681359142065048
step: 120, loss: 0.008782762102782726
step: 130, loss: 0.035619158297777176
step: 140, loss: 0.04142973944544792
step: 150, loss: 0.023188360035419464
step: 160, loss: 0.1661795824766159
step: 170, loss: 0.08231257647275925
step: 180, loss: 0.001487719127908349
step: 190, loss: 0.04796158894896507
step: 200, loss: 0.009560191072523594
step: 210, loss: 0.010697737336158752
/usr/local/lib/python3.7/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=0.7265774378585086, f1=0.72552783109405,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.045075200498104095
step: 10, loss: 0.020039869472384453
step: 20, loss: 0.015917303040623665
step: 30, loss: 0.025253286585211754
step: 40, loss: 0.03803640976548195
step: 50, loss: 0.10871313512325287
step: 60, loss: 0.1496177464723587

```

```

step: 70, loss: 0.0046210442669689655
step: 80, loss: 0.03571999445557594
step: 90, loss: 0.0027238831389695406
step: 100, loss: 0.012332464568316936
step: 110, loss: 0.05815236642956734
step: 120, loss: 0.026608647778630257
step: 130, loss: 0.002017971593886614
step: 140, loss: 0.014095094986259937
step: 150, loss: 0.04630356281995773
step: 160, loss: 0.014732099138200283
step: 170, loss: 0.028165897354483604
step: 180, loss: 0.06246339902281761
step: 190, loss: 0.005806421861052513
step: 200, loss: 0.008352760225534439
step: 210, loss: 0.017191648483276367
/usr/local/lib/python3.7/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=0.725868725868726, f1=0.7251908396946565,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00963661726564169
step: 10, loss: 0.004952209535986185
step: 20, loss: 0.010149484500288963
step: 30, loss: 0.030803892761468887
step: 40, loss: 0.029589645564556122
step: 50, loss: 0.01501902099698782
step: 60, loss: 0.08670095354318619
step: 70, loss: 0.003334689186885953
step: 80, loss: 0.0018417618703097105
step: 90, loss: 0.02423076704144478
step: 100, loss: 0.012032026425004005
step: 110, loss: 0.0033552385866642
step: 120, loss: 0.08800715953111649
step: 130, loss: 0.005048634018748999
step: 140, loss: 0.001043555443175137
step: 150, loss: 0.011853715404868126
step: 160, loss: 0.04274827241897583
step: 170, loss: 0.014262701384723186
step: 180, loss: 0.044689029455184937
step: 190, loss: 0.02363159880042076
step: 200, loss: 0.011422592215240002
step: 210, loss: 0.05554310977458954
/usr/local/lib/python3.7/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=0.7283236994219653, f1=0.7283464566929133,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009596267715096474
step: 10, loss: 0.04293445870280266
step: 20, loss: 0.04034390300512314
step: 30, loss: 0.07294320315122604
step: 40, loss: 0.09637010097503662
step: 50, loss: 0.0011020676465705037
step: 60, loss: 0.17031216621398926
step: 70, loss: 0.0026351038832217455
step: 80, loss: 0.016061263158917427
step: 90, loss: 0.09555377811193466
step: 100, loss: 0.05169657990336418
step: 110, loss: 0.018088659271597862
step: 120, loss: 0.010238922201097012
step: 130, loss: 0.0005453511839732528
step: 140, loss: 0.012070250697433949
step: 150, loss: 0.00267455424182117
step: 160, loss: 0.007293344475328922
step: 170, loss: 0.028348449617624283
step: 180, loss: 0.0694136694073677
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 190, loss: 0.12609544396400452
step: 200, loss: 0.011531557887792587
step: 210, loss: 0.018253464251756668
/usr/local/lib/python3.7/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=0.7265469061876247, f1=0.7134020618556701,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027489127591252327
step: 10, loss: 0.0007064450765028596
step: 20, loss: 0.01996488869190216
step: 30, loss: 0.006749078165739775
step: 40, loss: 0.05509147793054581
step: 50, loss: 0.07262715697288513
step: 60, loss: 0.0001936810731422156
step: 70, loss: 0.0010935742175206542
step: 80, loss: 0.00012248782149981707
step: 90, loss: 0.026460882276296616
step: 100, loss: 0.0001460371568100527
step: 110, loss: 0.0014669415540993214

```

```

step: 120, loss: 0.0073551912792027
step: 130, loss: 0.000981421209871769
step: 140, loss: 0.0018650790443643928
step: 150, loss: 0.0134569788351655
step: 160, loss: 0.0025606995914131403
step: 170, loss: 0.0345131978392601
step: 180, loss: 0.0007982732495293021
step: 190, loss: 0.04843754321336746
step: 200, loss: 0.006663412321358919
step: 210, loss: 0.0027628098614513874
/usr/local/lib/python3.7/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=0.7223300970873787, f1=0.7343750000000001,
best_f1=0.7347740667976425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009851028211414814
step: 10, loss: 0.03365310654044151
step: 20, loss: 0.0003464314213488251
step: 30, loss: 0.06929749995470047
step: 40, loss: 0.008558639325201511
step: 50, loss: 0.0023862833622843027
step: 60, loss: 0.0035495602060109377
step: 70, loss: 0.0014493323396891356
step: 80, loss: 0.060953740030527115
step: 90, loss: 0.0006441143923439085
step: 100, loss: 0.0005745688104070723
step: 110, loss: 0.0037470501847565174
step: 120, loss: 0.00019043294014409184
step: 130, loss: 0.026887349784374237
step: 140, loss: 0.017728393897414207
step: 150, loss: 0.0688738152384758
step: 160, loss: 0.0008136159740388393
step: 170, loss: 0.06630172580480576
step: 180, loss: 0.01532893255352974
step: 190, loss: 0.012984083034098148
step: 200, loss: 0.000179859678610228
step: 210, loss: 0.026900120079517365
/usr/local/lib/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=0.7291666666666666, f1=0.7208333333333333,

```

```

best_f1=0.7208333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028424616903066635
step: 10, loss: 0.0002372229064349085
step: 20, loss: 0.004309349227696657
step: 30, loss: 0.001023273915052414
step: 40, loss: 0.0017817316111177206
step: 50, loss: 0.0006896535051055253
step: 60, loss: 0.011400693096220493
step: 70, loss: 0.00023020636581350118
step: 80, loss: 0.0002939673140645027
step: 90, loss: 0.0010879153851419687
step: 100, loss: 0.0014545434387400746
step: 110, loss: 0.0007051757420413196
step: 120, loss: 0.007540376391261816
step: 130, loss: 0.01684407889842987
step: 140, loss: 0.0006156216841191053
step: 150, loss: 0.0001775055716279894
step: 160, loss: 0.0011040882673114538
step: 170, loss: 0.007992475293576717
step: 180, loss: 0.00028044998180121183
step: 190, loss: 0.13184626400470734
step: 200, loss: 0.02459080144762993
step: 210, loss: 0.003285832703113556
/usr/local/lib/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=0.742004264392324, f1=0.7086956521739131,
best_f1=0.7086956521739131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002817924541886896
step: 10, loss: 0.0018013479420915246
step: 20, loss: 0.0005029184394516051
step: 30, loss: 0.018759911879897118
step: 40, loss: 0.0002532737562432885
step: 50, loss: 0.0002734253357630223
step: 60, loss: 0.0009362722630612552
step: 70, loss: 0.013064492493867874
step: 80, loss: 0.0004286601906642318
step: 90, loss: 0.08142206817865372

```

```

step: 100, loss: 0.07049133628606796
step: 110, loss: 0.00018269868451170623
step: 120, loss: 0.00045406739809550345
step: 130, loss: 0.00011239809828111902
step: 140, loss: 0.01137803215533495
step: 150, loss: 0.006062960717827082
step: 160, loss: 0.03519753739237785
step: 170, loss: 0.0018413594225421548
step: 180, loss: 0.0006338239181786776
step: 190, loss: 0.0012234282912686467
step: 200, loss: 0.0006358250393532217
step: 210, loss: 0.00019577574857976288
/usr/local/lib/python3.7/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=0.7306122448979591, f1=0.7098121085594988,
best_f1=0.7086956521739131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005022529978305101
step: 10, loss: 0.0004544881812762469
step: 20, loss: 0.00024855328956618905
step: 30, loss: 0.00029530294705182314
step: 40, loss: 0.0008573928498663008
step: 50, loss: 0.0030126639176160097
step: 60, loss: 0.00353223760612309
step: 70, loss: 0.0016867663944140077
step: 80, loss: 0.06146950647234917
step: 90, loss: 0.00044958130456507206
step: 100, loss: 0.00030064801103435457
step: 110, loss: 0.0011404402321204543
step: 120, loss: 0.003964087925851345
step: 130, loss: 0.0002398611104581505
step: 140, loss: 0.0001197754536406137
step: 150, loss: 0.0005222646868787706
step: 160, loss: 0.0006074457196518779
step: 170, loss: 0.00015581068873871118
step: 180, loss: 0.006768071558326483
step: 190, loss: 0.0004643753345590085
step: 200, loss: 0.00025588006246834993
step: 210, loss: 0.002824086230248213
/usr/local/lib/python3.7/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=0.7215189873417722, f1=0.7078891257995735,
best_f1=0.7086956521739131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00803705956786871
step: 10, loss: 0.0005255148280411959
step: 20, loss: 0.00013677735114470124
step: 30, loss: 0.0002114178059855476
step: 40, loss: 0.00019486267410684377
step: 50, loss: 0.00011173862003488466
step: 60, loss: 0.004653358832001686
step: 70, loss: 0.0006060433224774897
step: 80, loss: 0.0002102848666254431
step: 90, loss: 0.002131071873009205
step: 100, loss: 8.376327605219558e-05
step: 110, loss: 7.374511187663302e-05
step: 120, loss: 0.00019670747860800475
step: 130, loss: 0.0012886004988104105
step: 140, loss: 0.0013709352351725101
step: 150, loss: 0.003786566201597452
step: 160, loss: 0.0005879667005501688
step: 170, loss: 0.019116004928946495
step: 180, loss: 0.08530773967504501
step: 190, loss: 0.00033813901245594025
step: 200, loss: 0.02093261480331421
step: 210, loss: 0.0051641990430653095
/usr/local/lib/python3.7/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=0.725, f1=0.7127882599580714, best_f1=0.7086956521739131
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, 345.12it/s]
load_f1 = 0.7209775967413441
real_f1 = 0.7099391480730223
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, 254.19it/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: 0.8489554524421692
```

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

```
step: 10, loss: 0.1650949865579605
```

```
step: 20, loss: 0.15148724615573883
```

```
step: 30, loss: 0.5069169402122498
```

```

step: 40, loss: 0.2526656985282898
step: 50, loss: 0.29979145526885986
step: 60, loss: 0.3589251637458801
step: 70, loss: 0.17978394031524658
step: 80, loss: 0.5278457403182983
step: 90, loss: 0.24259889125823975
step: 100, loss: 0.21885749697685242
step: 110, loss: 0.2381506860256195
step: 120, loss: 0.4051799476146698
step: 130, loss: 0.3399960696697235
step: 140, loss: 0.24244853854179382
step: 150, loss: 0.22609460353851318
step: 160, loss: 0.1839337944984436
step: 170, loss: 0.32175663113594055
step: 180, loss: 0.20632849633693695
step: 190, loss: 0.1551162600517273
/usr/local/lib/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=0.7157894736842104, f1=0.7045454545454547,
best_f1=0.7045454545454547
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1652025431394577
step: 10, loss: 0.025067593902349472
step: 20, loss: 0.07368715107440948
step: 30, loss: 0.10234536975622177
step: 40, loss: 0.34712520241737366
step: 50, loss: 0.1454273760318756
step: 60, loss: 0.2597101032733917
step: 70, loss: 0.15543575584888458
step: 80, loss: 0.1353473663330078
step: 90, loss: 0.21520572900772095
step: 100, loss: 0.37427884340286255
step: 110, loss: 0.11090636253356934
step: 120, loss: 0.11132003366947174
step: 130, loss: 0.16110345721244812
step: 140, loss: 0.09685564041137695
step: 150, loss: 0.021083291620016098
step: 160, loss: 0.06423033773899078
step: 170, loss: 0.11402961611747742
step: 180, loss: 0.13699832558631897
step: 190, loss: 0.09778101742267609

```

```

/usr/local/lib/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=0.7658402203856748, f1=0.8140161725067385,
best_f1=0.8140161725067385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07358285784721375
step: 10, loss: 0.16532571613788605
step: 20, loss: 0.023057837039232254
step: 30, loss: 0.02993893437087536
step: 40, loss: 0.011418310925364494
step: 50, loss: 0.11563337594270706
step: 60, loss: 0.05299065262079239
step: 70, loss: 0.15055213868618011
step: 80, loss: 0.08758491277694702
step: 90, loss: 0.02013937570154667
step: 100, loss: 0.050945643335580826
step: 110, loss: 0.09018465131521225
step: 120, loss: 0.06675630807876587
step: 130, loss: 0.007822184823453426
step: 140, loss: 0.027397099882364273
step: 150, loss: 0.07528757303953171
step: 160, loss: 0.2747235596179962
step: 170, loss: 0.03160262852907181
step: 180, loss: 0.09605880081653595
step: 190, loss: 0.06587722897529602
/usr/local/lib/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=0.7683923705722071, f1=0.7520891364902507,
best_f1=0.7520891364902507
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02057591639459133
step: 10, loss: 0.013665856793522835
step: 20, loss: 0.009094675071537495
step: 30, loss: 0.019890204071998596

```

```

step: 40, loss: 0.017356326803565025
step: 50, loss: 0.005322582554072142
step: 60, loss: 0.02973249740898609
step: 70, loss: 0.0040740882977843285
step: 80, loss: 0.11311502009630203
step: 90, loss: 0.003933212719857693
step: 100, loss: 0.016978120431303978
step: 110, loss: 0.0014546310994774103
step: 120, loss: 0.03700350597500801
step: 130, loss: 0.13750863075256348
step: 140, loss: 0.08357876539230347
step: 150, loss: 0.04178842157125473
step: 160, loss: 0.02265012450516224
step: 170, loss: 0.008092127740383148
step: 180, loss: 0.15392617881298065
step: 190, loss: 0.03933577239513397
/usr/local/lib/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=0.7806122448979592, f1=0.7700534759358288,
best_f1=0.7700534759358288
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06993557512760162
step: 10, loss: 0.014088747091591358
step: 20, loss: 0.049512363970279694
step: 30, loss: 0.10011307150125504
step: 40, loss: 0.03603951632976532
step: 50, loss: 0.01727273128926754
step: 60, loss: 0.005595670081675053
step: 70, loss: 0.0029358412139117718
step: 80, loss: 0.0015825387090444565
step: 90, loss: 0.0065209814347326756
step: 100, loss: 0.11008137464523315
step: 110, loss: 0.0005148960626684129
step: 120, loss: 0.0012246391270309687
step: 130, loss: 0.04960361495614052
step: 140, loss: 0.012232690118253231
step: 150, loss: 0.019452335312962532
step: 160, loss: 0.0025744892191141844
step: 170, loss: 0.10637672990560532
step: 180, loss: 0.10301622003316879
step: 190, loss: 0.04183650016784668

```

```

/usr/local/lib/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=0.7859078590785908, f1=0.7754010695187165,
best_f1=0.7754010695187165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003718622261658311
step: 10, loss: 0.005517437122762203
step: 20, loss: 0.0013957235496491194
step: 30, loss: 0.021666839718818665
step: 40, loss: 0.002033126074820757
step: 50, loss: 0.014253895729780197
step: 60, loss: 0.0015627279644832015
step: 70, loss: 0.014577574096620083
step: 80, loss: 0.027178416028618813
step: 90, loss: 0.02530910074710846
step: 100, loss: 0.01595431938767433
step: 110, loss: 0.00197125063277781
step: 120, loss: 0.00255908933468163
step: 130, loss: 0.001884067663922906
step: 140, loss: 0.0009716291096992791
step: 150, loss: 0.0050200242549180984
step: 160, loss: 0.024515049532055855
step: 170, loss: 0.026430441066622734
step: 180, loss: 0.017740609124302864
step: 190, loss: 0.003433778416365385
/usr/local/lib/python3.7/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=0.78125, f1=0.7783783783783784, best_f1=0.7754010695187165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038579225074499846
step: 10, loss: 0.005461592692881823
step: 20, loss: 0.002128738909959793
step: 30, loss: 0.003061917843297124
step: 40, loss: 0.006063135806471109
step: 50, loss: 0.018351376056671143
step: 60, loss: 0.1071481853723526
step: 70, loss: 0.005808877758681774
step: 80, loss: 0.012576278299093246

```

```

step: 90, loss: 0.02241169847548008
step: 100, loss: 0.0022494622971862555
step: 110, loss: 0.0014356314204633236
step: 120, loss: 0.027165329083800316
step: 130, loss: 0.001231830450706184
step: 140, loss: 0.00037376812542788684
step: 150, loss: 0.026306072250008583
step: 160, loss: 0.0005340245552361012
step: 170, loss: 0.0024255504831671715
step: 180, loss: 0.0019057802855968475
step: 190, loss: 0.0007599266245961189
/usr/local/lib/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=0.7914438502673796, f1=0.7799442896935933,
best_f1=0.7799442896935933
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005140276625752449
step: 10, loss: 0.006148581858724356
step: 20, loss: 0.0008156165713444352
step: 30, loss: 0.0009628896950744092
step: 40, loss: 0.0008317145402543247
step: 50, loss: 0.0016668298048898578
step: 60, loss: 0.0023095125798135996
step: 70, loss: 0.000864940055180341
step: 80, loss: 0.09341160207986832
step: 90, loss: 0.0009228636627085507
step: 100, loss: 0.048598628491163254
step: 110, loss: 0.0010896301828324795
step: 120, loss: 0.008252057246863842
step: 130, loss: 0.008787417784333229
step: 140, loss: 0.006936237215995789
step: 150, loss: 0.0008871473255567253
step: 160, loss: 0.00041971320752054453
step: 170, loss: 0.00033410420292057097
step: 180, loss: 0.12802112102508545
step: 190, loss: 0.004454801790416241
/usr/local/lib/python3.7/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=0.774025974025974, f1=0.7893333333333333,
best_f1=0.7799442896935933

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026241453364491463
step: 10, loss: 0.014179486781358719
step: 20, loss: 0.004515079781413078
step: 30, loss: 0.0006837717955932021
step: 40, loss: 0.003163010347634554
step: 50, loss: 0.003892812877893448
step: 60, loss: 0.0011500604450702667
step: 70, loss: 0.0004057384794577956
step: 80, loss: 0.00039057427784428
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.004686895292252302
step: 100, loss: 0.00036266256938688457
step: 110, loss: 0.0033080540597438812
step: 120, loss: 0.0017579930135980248
step: 130, loss: 0.0003775829100050032
step: 140, loss: 0.00023756206792313606
step: 150, loss: 0.0023195031099021435
step: 160, loss: 0.0021012856159359217
step: 170, loss: 0.00048177503049373627
step: 180, loss: 0.00027073570527136326
step: 190, loss: 0.11131465435028076
/usr/local/lib/python3.7/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=0.7862796833773087, f1=0.7957559681697614,
best_f1=0.7799442896935933
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013375081471167505
step: 10, loss: 0.00026850911672227085
step: 20, loss: 0.000832683639600873
step: 30, loss: 0.06183455511927605
step: 40, loss: 0.0030437177047133446
step: 50, loss: 0.0001686844479991123
step: 60, loss: 0.010475309565663338
step: 70, loss: 0.010005120187997818
step: 80, loss: 0.0011652734829112887
step: 90, loss: 0.001836627721786499
step: 100, loss: 0.002674718853086233
step: 110, loss: 0.0035028401762247086
step: 120, loss: 0.01101684756577015
step: 130, loss: 0.0016355168772861362
step: 140, loss: 0.0033675185404717922
step: 150, loss: 0.0006025838665664196

```



```

step: 160, loss: 0.001223031897097826
step: 170, loss: 0.0005325382226146758
step: 180, loss: 0.0007856519659981132
step: 190, loss: 0.001771777169778943
/usr/local/lib/python3.7/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=0.7783505154639175, f1=0.7853403141361257,
best_f1=0.7799442896935933
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00024394405772909522
step: 10, loss: 0.0006808128673583269
step: 20, loss: 0.0007498895865865052
step: 30, loss: 0.019342102110385895
step: 40, loss: 0.00022807173081673682
step: 50, loss: 0.0006218741764314473
step: 60, loss: 0.07211371511220932
step: 70, loss: 0.00017117692914325744
step: 80, loss: 0.00015018758131191134
step: 90, loss: 0.0006933336844667792
step: 100, loss: 0.0002649320231284946
step: 110, loss: 0.0003897145506925881
step: 120, loss: 0.0011495943181216717
step: 130, loss: 0.0004154528141953051
step: 140, loss: 0.0001811760157579556
step: 150, loss: 0.006093223579227924
step: 160, loss: 0.0003583697834983468
step: 170, loss: 0.0045720795169472694
step: 180, loss: 0.0232230331748724
step: 190, loss: 0.0005507540190592408
/usr/local/lib/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=0.7917737789203083, f1=0.7897435897435896,
best_f1=0.7897435897435896
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005021069082431495
step: 10, loss: 0.001413369900546968
step: 20, loss: 0.00035355688305571675
step: 30, loss: 0.018250171095132828

```

```

step: 40, loss: 0.0008850686135701835
step: 50, loss: 0.0005385233671404421
step: 60, loss: 0.0003134622238576412
step: 70, loss: 0.00046700809616595507
step: 80, loss: 0.00042177733848802745
step: 90, loss: 0.00042097666300833225
step: 100, loss: 0.001730639487504959
step: 110, loss: 0.0004226279561407864
step: 120, loss: 0.0010202578268945217
step: 130, loss: 0.0006536630680784583
step: 140, loss: 0.0007412632112391293
step: 150, loss: 0.0003098263405263424
step: 160, loss: 0.00017275824211537838
step: 170, loss: 0.05085724592208862
step: 180, loss: 0.00042965624015778303
step: 190, loss: 0.007195806130766869
/usr/local/lib/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=0.7949999999999999, f1=0.7938144329896906,
best_f1=0.7938144329896906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004856355953961611
step: 10, loss: 0.0007147466531023383
step: 20, loss: 0.00032221703440882266
step: 30, loss: 0.00016341981245204806
step: 40, loss: 0.00034113178844563663
step: 50, loss: 0.0011557817924767733
step: 60, loss: 0.00035595204099081457
step: 70, loss: 0.000303929322399199
step: 80, loss: 0.00022354583779815584
step: 90, loss: 0.00020287504594307393
step: 100, loss: 0.0007516528130508959
step: 110, loss: 0.00023718264128547162
step: 120, loss: 0.0060859196819365025
step: 130, loss: 0.0008094244985841215
step: 140, loss: 0.000584171328227967
step: 150, loss: 0.0014853216707706451
step: 160, loss: 0.00027694282471202314
step: 170, loss: 0.0004733246751129627
step: 180, loss: 0.0072050923481583595
step: 190, loss: 0.0005767035181634128

```

```

/usr/local/lib/python3.7/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=0.7868020304568528, f1=0.7874015748031497,
best_f1=0.7938144329896906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02815057337284088
step: 10, loss: 0.0007948774145916104
step: 20, loss: 0.0003804876178037375
step: 30, loss: 0.006339015904814005
step: 40, loss: 0.0002707800595089793
step: 50, loss: 0.00017745727382134646
step: 60, loss: 0.00012933067046105862
step: 70, loss: 0.00109420670196414
step: 80, loss: 0.00010468435357324779
step: 90, loss: 0.0015026003820821643
step: 100, loss: 0.003933769650757313
step: 110, loss: 0.0002349006972508505
step: 120, loss: 0.002188359387218952
step: 130, loss: 0.00012244134268257767
step: 140, loss: 0.0017622180748730898
step: 150, loss: 0.00020877690985798836
step: 160, loss: 0.00021141083561815321
step: 170, loss: 0.0003722903202287853
step: 180, loss: 0.00020335291628725827
step: 190, loss: 0.00031054342980496585
/usr/local/lib/python3.7/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=0.7938144329896906, f1=0.7894736842105263,
best_f1=0.7938144329896906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00027281453367322683
step: 10, loss: 0.00016528363630641252
step: 20, loss: 0.000386317289667204
step: 30, loss: 0.000920614693313837
step: 40, loss: 0.0002847537980414927
step: 50, loss: 0.0005582338199019432
step: 60, loss: 0.0002069263136945665
step: 70, loss: 0.0005048710736446083
step: 80, loss: 0.0010471352143213153
step: 90, loss: 0.001705571310594678
step: 100, loss: 0.013014478608965874
step: 110, loss: 0.0001453275908716023

```

```

step: 120, loss: 0.0003387547330930829
step: 130, loss: 0.00029404074302874506
step: 140, loss: 0.0012841591378673911
step: 150, loss: 0.00012425209570210427
step: 160, loss: 0.001451305579394102
step: 170, loss: 0.00016972511366475374
step: 180, loss: 0.0005014636553823948
step: 190, loss: 0.0018973794067278504
/usr/local/lib/python3.7/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=0.7857142857142857, f1=0.7894736842105263,
best_f1=0.7938144329896906
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, 214.81it/s]
load_f1 = 0.6244541484716156
real_f1 = 0.6101694915254237
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, 257.16it/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: 0.8520340919494629
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22488844394683838
step: 20, loss: 0.15698660910129547
step: 30, loss: 0.22186067700386047
step: 40, loss: 0.3174856901168823
step: 50, loss: 0.3774935007095337
step: 60, loss: 0.4449770748615265
step: 70, loss: 0.31368857622146606
step: 80, loss: 0.2514837086200714
step: 90, loss: 0.4010474681854248
step: 100, loss: 0.22802609205245972
step: 110, loss: 0.16737599670886993
step: 120, loss: 0.5246094465255737
step: 130, loss: 0.3816835284233093
step: 140, loss: 0.37378156185150146
step: 150, loss: 0.13974258303642273
step: 160, loss: 0.15440872311592102
step: 170, loss: 0.1130031943321228

```

```

/usr/local/lib/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=0.7289293849658314, f1=0.7164835164835165,
best_f1=0.7164835164835165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20326505601406097
step: 10, loss: 0.07123638689517975
step: 20, loss: 0.18322081863880157
step: 30, loss: 0.08834145218133926
step: 40, loss: 0.176923006772995
step: 50, loss: 0.19133013486862183
step: 60, loss: 0.05375029146671295
step: 70, loss: 0.1196746677160263
step: 80, loss: 0.1460787057876587
step: 90, loss: 0.15653438866138458
step: 100, loss: 0.16083835065364838
step: 110, loss: 0.15282486379146576
step: 120, loss: 0.10268960893154144
step: 130, loss: 0.07596711814403534
step: 140, loss: 0.10301432758569717
step: 150, loss: 0.05176176130771637
step: 160, loss: 0.08825948089361191
step: 170, loss: 0.09590191394090652
/usr/local/lib/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=0.7989690721649484, f1=0.8028846153846153,
best_f1=0.8028846153846153
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022606102749705315
step: 10, loss: 0.035992320626974106
step: 20, loss: 0.05120282247662544
step: 30, loss: 0.17329803109169006
step: 40, loss: 0.01560852862894535
step: 50, loss: 0.047808002680540085

```

```

step: 60, loss: 0.06190841645002365
step: 70, loss: 0.08533938974142075
step: 80, loss: 0.05502813309431076
step: 90, loss: 0.19667018949985504
step: 100, loss: 0.013058903627097607
step: 110, loss: 0.10903346538543701
step: 120, loss: 0.00500030443072319
step: 130, loss: 0.14384053647518158
step: 140, loss: 0.0034169272985309362
step: 150, loss: 0.06110804155468941
step: 160, loss: 0.03028026781976223
step: 170, loss: 0.0879216119647026
/usr/local/lib/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=0.8112244897959183, f1=0.7980997624703088,
best_f1=0.7980997624703088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05058663338422775
step: 10, loss: 0.026450825855135918
step: 20, loss: 0.030065573751926422
step: 30, loss: 0.03819425031542778
step: 40, loss: 0.005692628212273121
step: 50, loss: 0.009374348446726799
step: 60, loss: 0.0515114963054657
step: 70, loss: 0.003325848374515772
step: 80, loss: 0.07684038579463959
step: 90, loss: 0.014728515408933163
step: 100, loss: 0.019947366788983345
step: 110, loss: 0.02895507961511612
step: 120, loss: 0.11153930425643921
step: 130, loss: 0.011360005475580692
step: 140, loss: 0.010418349876999855
step: 150, loss: 0.002011645818129182
step: 160, loss: 0.03875793144106865
step: 170, loss: 0.008120572194457054
/usr/local/lib/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=0.8395061728395062, f1=0.8120649651972157,
best_f1=0.8120649651972157
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005757858976721764
step: 10, loss: 0.05832966789603233
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 20, loss: 0.011649494990706444
step: 30, loss: 0.03995661437511444
step: 40, loss: 0.003880159230902791
step: 50, loss: 0.002765523036941886
step: 60, loss: 0.001648825011216104
step: 70, loss: 0.13710199296474457
step: 80, loss: 0.008399990387260914
step: 90, loss: 0.022904422134160995
step: 100, loss: 0.04257849603891373
step: 110, loss: 0.18681137263774872
step: 120, loss: 0.05100514739751816
step: 130, loss: 0.0024302322417497635
step: 140, loss: 0.019986050203442574
step: 150, loss: 0.008758928626775742
step: 160, loss: 0.003439917927607894
step: 170, loss: 0.029033254832029343
/usr/local/lib/python3.7/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=0.8345323741007193, f1=0.8154897494305238,
best_f1=0.8120649651972157
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03884810209274292
step: 10, loss: 0.07593899220228195
step: 20, loss: 0.011659125797450542
step: 30, loss: 0.0043466477654874325
step: 40, loss: 0.001860840362496674
step: 50, loss: 0.0012237231712788343
step: 60, loss: 0.11870995908975601
step: 70, loss: 0.008390911854803562
step: 80, loss: 0.014766168780624866
step: 90, loss: 0.02513655461370945
step: 100, loss: 0.013644675724208355
step: 110, loss: 0.00522536551579833
step: 120, loss: 0.03051506169140339
step: 130, loss: 0.16708873212337494
step: 140, loss: 0.020120173692703247

```

```

step: 150, loss: 0.10252472758293152
step: 160, loss: 0.10246175527572632
step: 170, loss: 0.005153097677975893
/usr/local/lib/python3.7/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=0.8353221957040573, f1=0.8183908045977012,
best_f1=0.8120649651972157
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0035835979506373405
step: 10, loss: 0.0006203849916346371
step: 20, loss: 0.010305207222700119
step: 30, loss: 0.0071134683676064014
step: 40, loss: 0.02286265417933464
step: 50, loss: 0.0029339164029806852
step: 60, loss: 0.09675188362598419
step: 70, loss: 0.0054527390748262405
step: 80, loss: 0.006314470898360014
step: 90, loss: 0.029775459319353104
step: 100, loss: 0.037407439202070236
step: 110, loss: 0.0011989010963588953
step: 120, loss: 0.12463854253292084
step: 130, loss: 0.018595481291413307
step: 140, loss: 0.02093677595257759
step: 150, loss: 0.03438088297843933
step: 160, loss: 0.010890456847846508
step: 170, loss: 0.08772888779640198
/usr/local/lib/python3.7/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=0.8372093023255814, f1=0.8162291169451074,
best_f1=0.8120649651972157
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03254402428865433
step: 10, loss: 0.0008256776491180062
step: 20, loss: 0.005625675432384014
step: 30, loss: 0.0031712185591459274
step: 40, loss: 0.002699791919440031
step: 50, loss: 0.0002807457931339741
step: 60, loss: 0.0024479886051267385
step: 70, loss: 0.021758442744612694
step: 80, loss: 0.03408004716038704
step: 90, loss: 0.00672830268740654
step: 100, loss: 0.021106574684381485

```

```

step: 110, loss: 0.09628456085920334
step: 120, loss: 0.00039007433224469423
step: 130, loss: 0.00047122902469709516
step: 140, loss: 0.0517059788107872
step: 150, loss: 0.04027504473924637
step: 160, loss: 0.013944401405751705
step: 170, loss: 0.012740105390548706
/usr/local/lib/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=0.8544600938967137, f1=0.8008948545861299,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013149079168215394
step: 10, loss: 0.009363159537315369
step: 20, loss: 0.09265786409378052
step: 30, loss: 0.09344994276762009
step: 40, loss: 0.01880464516580105
step: 50, loss: 0.014844097197055817
step: 60, loss: 0.001083719776943326
step: 70, loss: 0.002000270178541541
step: 80, loss: 0.04257944971323013
step: 90, loss: 0.002936603268608451
step: 100, loss: 0.030698716640472412
step: 110, loss: 8.059270476223901e-05
step: 120, loss: 0.0034948387183248997
step: 130, loss: 0.006742862053215504
step: 140, loss: 0.0023954990319907665
step: 150, loss: 0.01553534995764494
step: 160, loss: 0.0670761913061142
step: 170, loss: 0.03319338336586952
/usr/local/lib/python3.7/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=0.8383838383838383, f1=0.8300970873786409,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004951125010848045
step: 10, loss: 0.016131525859236717
step: 20, loss: 0.17514581978321075

```

```

step: 30, loss: 0.0001750906085362658
step: 40, loss: 0.001614622538909316
step: 50, loss: 0.0026275969576090574
step: 60, loss: 0.001376432366669178
step: 70, loss: 0.006503427401185036
step: 80, loss: 0.0036426254082471132
step: 90, loss: 0.02796691469848156
step: 100, loss: 0.0028861805330961943
step: 110, loss: 0.0004112664901185781
step: 120, loss: 0.02532140724360943
step: 130, loss: 0.013155057094991207
step: 140, loss: 0.014531160704791546
step: 150, loss: 0.0004264705639798194
step: 160, loss: 0.012822183780372143
step: 170, loss: 0.0011789381969720125
/usr/local/lib/python3.7/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=0.8255528255528256, f1=0.8141176470588235,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000626903201919049
step: 10, loss: 0.0011620864970609546
step: 20, loss: 0.04708482325077057
step: 30, loss: 0.0007970881997607648
step: 40, loss: 0.000971776491496712
step: 50, loss: 0.0008353866287507117
step: 60, loss: 0.0018593573477119207
step: 70, loss: 0.000566434464417398
step: 80, loss: 0.00035401206696406007
step: 90, loss: 0.005408828612416983
step: 100, loss: 0.00011759463814087212
step: 110, loss: 0.0019362557213753462
step: 120, loss: 0.007703141309320927
step: 130, loss: 0.00444705318659544
step: 140, loss: 0.026510247960686684
step: 150, loss: 0.003385928925126791
step: 160, loss: 0.006304811220616102
step: 170, loss: 0.049036357551813126
/usr/local/lib/python3.7/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=0.8522167487684729, f1=0.8249400479616307,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01837899535894394
step: 10, loss: 0.0037985765375196934
step: 20, loss: 0.024993373081088066
step: 30, loss: 0.0036937615368515253
step: 40, loss: 0.00036887705209665
step: 50, loss: 0.0002055898221442476
step: 60, loss: 0.0008429419831372797
step: 70, loss: 0.007954582571983337
step: 80, loss: 0.01559496484696865
step: 90, loss: 0.0006875915569253266
step: 100, loss: 0.018365276977419853
step: 110, loss: 0.0001396315055899322
step: 120, loss: 0.0006573798018507659
step: 130, loss: 0.0046107470989227295
step: 140, loss: 0.00026667918427847326
step: 150, loss: 0.0003466703637968749
step: 160, loss: 0.0027571695391088724
step: 170, loss: 0.001143097411841154
/usr/local/lib/python3.7/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=0.8502415458937198, f1=0.8221709006928406,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06351940333843231
step: 10, loss: 0.04848160222172737
step: 20, loss: 0.000573648139834404
step: 30, loss: 0.0032650348730385303
step: 40, loss: 0.0002851342724170536
step: 50, loss: 0.0009057588176801801
step: 60, loss: 0.00026465801056474447
step: 70, loss: 0.008572924882173538
step: 80, loss: 0.0005119367269799113
step: 90, loss: 0.0002271074044983834
step: 100, loss: 0.00019052538846153766
step: 110, loss: 0.0004798125592060387
step: 120, loss: 0.0004093461611773819
step: 130, loss: 0.004022911656647921
step: 140, loss: 0.0004415092698764056
step: 150, loss: 0.00020476552890613675
step: 160, loss: 0.0015307458816096187
step: 170, loss: 0.001159332925453782
/usr/local/lib/python3.7/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=0.8535980148883374, f1=0.830188679245283,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00040445211925543845
step: 10, loss: 0.0017575842794030905
step: 20, loss: 0.011887640692293644
step: 30, loss: 0.00834919884800911
step: 40, loss: 0.00045402260730043054
step: 50, loss: 0.0009198523475788534
step: 60, loss: 0.002049177186563611
step: 70, loss: 0.008824566379189491
step: 80, loss: 0.001055508735589683
step: 90, loss: 0.00019891001284122467
step: 100, loss: 0.0001417541498085484
step: 110, loss: 0.00022669366444461048
step: 120, loss: 0.11099052429199219
step: 130, loss: 0.00042972349910996854
step: 140, loss: 0.0006676894845440984
step: 150, loss: 0.0004461452772375196
step: 160, loss: 0.01668139174580574
step: 170, loss: 0.00017050014866981655
/usr/local/lib/python3.7/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=0.848780487804878, f1=0.8205128205128204,
best_f1=0.8008948545861299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002437883522361517
step: 10, loss: 0.014364118687808514
step: 20, loss: 0.018740219995379448
step: 30, loss: 0.04517466202378273
step: 40, loss: 0.0001649031473789364
step: 50, loss: 0.010249794460833073
step: 60, loss: 8.442264515906572e-05
step: 70, loss: 0.0050888461992144585
step: 80, loss: 0.028750978410243988
step: 90, loss: 0.0006732146139256656
step: 100, loss: 0.00046880761510692537
step: 110, loss: 0.00013551472511608154
step: 120, loss: 0.007791923359036446
step: 130, loss: 0.0036476394161581993
step: 140, loss: 0.0001697084226179868
step: 150, loss: 0.043543536216020584
step: 160, loss: 0.00011142114090034738

```

```

step: 170, loss: 0.00025515982997603714
/usr/local/lib/python3.7/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=0.8446601941747571, f1=0.8148148148148148,
best_f1=0.8008948545861299
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.")
1916it [00:05, 325.03it/s]
load_f1 = 0.4714038128249567
real_f1 = 0.45614035087719307
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, 253.20it/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: 0.8003515005111694
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45904645323753357
step: 20, loss: 0.5699005126953125
step: 30, loss: 0.3981249928474426
step: 40, loss: 0.1451943814754486
step: 50, loss: 0.049478691071271896
step: 60, loss: 0.12806840240955353
step: 70, loss: 0.05042334645986557
step: 80, loss: 0.23008304834365845
step: 90, loss: 0.09240416437387466
step: 100, loss: 0.3364010453224182
step: 110, loss: 0.13307543098926544
step: 120, loss: 0.0337943360209465
step: 130, loss: 0.005816265475004911
step: 140, loss: 0.004905201960355043
step: 150, loss: 0.1512766182422638
step: 160, loss: 0.20285388827323914
step: 170, loss: 0.009923257865011692
step: 180, loss: 0.023350656032562256
step: 190, loss: 0.06400842219591141
step: 200, loss: 0.004814969841390848
step: 210, loss: 0.021740084514021873

```

```

step: 220, loss: 0.007476418279111385
step: 230, loss: 0.017863694578409195
/usr/local/lib/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=0.9646017699115044, f1=0.9655172413793103,
best_f1=0.9655172413793103
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009619745425879955
step: 10, loss: 0.0009299519006162882
step: 20, loss: 0.0018804221181198955
step: 30, loss: 0.0015147054800763726
step: 40, loss: 0.00758657930418849
step: 50, loss: 0.023115474730730057
step: 60, loss: 0.00988516304641962
step: 70, loss: 0.07083023339509964
step: 80, loss: 0.0023394725285470486
step: 90, loss: 0.00753247132524848
step: 100, loss: 0.15009167790412903
step: 110, loss: 0.10318923741579056
step: 120, loss: 0.006281467620283365
step: 130, loss: 0.008773798123002052
step: 140, loss: 0.3176230192184448
step: 150, loss: 0.016801591962575912
step: 160, loss: 0.01700682006776333
step: 170, loss: 0.00458877719938755
step: 180, loss: 0.008797773160040379
step: 190, loss: 0.08973225206136703
step: 200, loss: 0.00845953356474638
step: 210, loss: 0.05566846579313278
step: 220, loss: 0.00377077073790133
step: 230, loss: 0.005837474018335342
/usr/local/lib/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=0.9744160177975528, f1=0.970917225950783,
best_f1=0.970917225950783
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09002438932657242
step: 10, loss: 0.009816575795412064
step: 20, loss: 0.005078346002846956
step: 30, loss: 0.004528108052909374
step: 40, loss: 0.041037216782569885
step: 50, loss: 0.02000393718481064
step: 60, loss: 0.0011189311044290662
step: 70, loss: 0.0016380584565922618
step: 80, loss: 0.0011182427406311035
step: 90, loss: 0.1493077129125595
step: 100, loss: 0.003204992739483714
step: 110, loss: 0.0033022966235876083
step: 120, loss: 0.012141261249780655
step: 130, loss: 0.0020868340507149696
step: 140, loss: 0.004671670962125063
step: 150, loss: 0.0030612857080996037
step: 160, loss: 0.006526478566229343
step: 170, loss: 0.00454039266332984
step: 180, loss: 0.0022870099637657404
step: 190, loss: 0.08647189289331436
step: 200, loss: 0.008653655648231506
step: 210, loss: 0.023015255108475685
step: 220, loss: 0.011354741640388966
step: 230, loss: 0.04874582216143608
/usr/local/lib/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=0.9776785714285714, f1=0.9755011135857461,
best_f1=0.9755011135857461
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018405260518193245
step: 10, loss: 0.00217633368447423
step: 20, loss: 0.0004735120164696127
step: 30, loss: 0.0009512454271316528
step: 40, loss: 0.006233533378690481
step: 50, loss: 0.0036556716077029705
step: 60, loss: 0.0023834628518670797
step: 70, loss: 0.008355548605322838
step: 80, loss: 0.1269964575767517
step: 90, loss: 0.003995761275291443

```

```

step: 100, loss: 0.0019355149706825614
step: 110, loss: 0.011204471811652184
step: 120, loss: 0.008769242092967033
step: 130, loss: 0.013783520087599754
step: 140, loss: 0.0006042806198820472
step: 150, loss: 0.0013434740249067545
step: 160, loss: 0.002776224398985505
step: 170, loss: 0.005314216483384371
step: 180, loss: 0.11047349125146866
step: 190, loss: 0.004966201260685921
step: 200, loss: 0.005379219073802233
step: 210, loss: 0.002556499792262912
step: 220, loss: 0.0027822619304060936
step: 230, loss: 0.0004112937895115465
/usr/local/lib/python3.7/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=0.9678848283499446, f1=0.9733333333333333,
best_f1=0.9755011135857461
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002118996111676097
step: 10, loss: 0.0015447678742930293
step: 20, loss: 0.0009572518174536526
step: 30, loss: 0.00034256631624884903
step: 40, loss: 0.0007000101031735539
step: 50, loss: 0.0002705338702071458
step: 60, loss: 0.0007032481953501701
step: 70, loss: 0.00020922870317008346
step: 80, loss: 0.0006601616041734815
step: 90, loss: 0.013818993233144283
step: 100, loss: 0.002585629466921091
step: 110, loss: 0.0008511510095559061
step: 120, loss: 0.022204121574759483
step: 130, loss: 0.10506503283977509
step: 140, loss: 0.0014422811800614
step: 150, loss: 0.00026541503029875457
step: 160, loss: 0.10881850123405457
step: 170, loss: 0.03989902883768082
step: 180, loss: 0.0010073411976918578
step: 190, loss: 0.0007456658640876412
step: 200, loss: 0.005016316194087267
step: 210, loss: 0.007793172262609005
step: 220, loss: 0.0018631925340741873
step: 230, loss: 0.006601481698453426
/usr/local/lib/python3.7/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=0.9733924611973392, f1=0.9744160177975528,
best_f1=0.9755011135857461
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038654087111353874
step: 10, loss: 0.0004667936882469803
step: 20, loss: 0.0011225466150790453
step: 30, loss: 0.0018428784096613526
step: 40, loss: 0.0010657422244548798
step: 50, loss: 0.14744439721107483
step: 60, loss: 0.006568963639438152
step: 70, loss: 0.0008619080181233585
step: 80, loss: 0.002327114576473832
step: 90, loss: 0.0006896751001477242
step: 100, loss: 0.0007571071037091315
step: 110, loss: 0.1394466757774353
step: 120, loss: 0.0004356068093329668
step: 130, loss: 0.006979909725487232
step: 140, loss: 0.0013061274075880647
step: 150, loss: 0.00048421008978039026
step: 160, loss: 0.0006983732455410063
step: 170, loss: 0.0042530507780611515
step: 180, loss: 0.0036094775423407555
step: 190, loss: 0.038202326744794846
step: 200, loss: 0.0020535888615995646
step: 210, loss: 0.0031151622533798218
step: 220, loss: 0.00029022074886597693
step: 230, loss: 0.0009676578338257968
/usr/local/lib/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=0.977924944812362, f1=0.9734513274336283,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09924639016389847
step: 10, loss: 0.0003045535704586655
step: 20, loss: 0.00029113495838828385
step: 30, loss: 0.00043301889672875404
step: 40, loss: 0.0025472831912338734
step: 50, loss: 0.00022307870676741004

```

```

step: 60, loss: 0.014868393540382385
step: 70, loss: 0.0005941253621131182
step: 80, loss: 0.0004996447241865098
step: 90, loss: 0.0010412951232865453
step: 100, loss: 0.008224420249462128
step: 110, loss: 0.0046218279749155045
step: 120, loss: 0.004114252980798483
step: 130, loss: 0.003137500723823905
step: 140, loss: 0.00014805216051172465
step: 150, loss: 0.004265293013304472
step: 160, loss: 0.0003920308081433177
step: 170, loss: 0.00043609869317151606
step: 180, loss: 0.0004593084449879825
step: 190, loss: 0.0003823166189249605
step: 200, loss: 0.0009282040991820395
step: 210, loss: 0.00017596743418835104
step: 220, loss: 0.0003168834373354912
step: 230, loss: 0.02667456679046154
/usr/local/lib/python3.7/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=0.9754464285714286, f1=0.9754464285714286,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009918435476720333
step: 10, loss: 0.008744274266064167
step: 20, loss: 0.00018542460747994483
step: 30, loss: 0.0004996635834686458
step: 40, loss: 0.0019222864648327231
step: 50, loss: 0.0003022572200279683
step: 60, loss: 0.00021356080833356827
step: 70, loss: 0.00012157829769421369
step: 80, loss: 9.874218085315078e-05
step: 90, loss: 9.878147102426738e-05
step: 100, loss: 0.00017159547132905573
step: 110, loss: 0.0001013376604532823
step: 120, loss: 5.8909568906528875e-05
step: 130, loss: 0.0004372352850623429
step: 140, loss: 0.00014176576223690063
step: 150, loss: 6.197137554408982e-05
step: 160, loss: 0.00041128715383820236
step: 170, loss: 0.00029826577519997954
step: 180, loss: 0.0005585555336438119
step: 190, loss: 4.73014879389666e-05
step: 200, loss: 0.0014169147470965981
step: 210, loss: 0.00010518771887291223

```

```

step: 220, loss: 9.471082012169063e-05
step: 230, loss: 0.06281889975070953
/usr/local/lib/python3.7/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=0.9765886287625419, f1=0.9700996677740864,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001619700633455068
step: 10, loss: 8.787489787209779e-05
step: 20, loss: 0.0034925779327750206
step: 30, loss: 0.0024911833461374044
step: 40, loss: 6.814180960645899e-05
step: 50, loss: 0.00015170486585702747
step: 60, loss: 0.00016420084284618497
step: 70, loss: 0.0005956473760306835
step: 80, loss: 0.06971774250268936
step: 90, loss: 0.009596574120223522
step: 100, loss: 0.01015542633831501
step: 110, loss: 7.489261770388111e-05
step: 120, loss: 0.05019181966781616
step: 130, loss: 7.728965283604339e-05
step: 140, loss: 0.008174656890332699
step: 150, loss: 5.896322909393348e-05
step: 160, loss: 0.00056826276704669
step: 170, loss: 0.00013851586845703423
step: 180, loss: 0.00014847778948023915
step: 190, loss: 0.009079021401703358
step: 200, loss: 0.010219878517091274
step: 210, loss: 0.09007580578327179
step: 220, loss: 0.03787285462021828
step: 230, loss: 0.0997389703989029
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.9776785714285714,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037203289684839547
step: 10, loss: 0.00027320985100232065
step: 20, loss: 0.012139207683503628
step: 30, loss: 0.0003608506522141397
step: 40, loss: 0.0008538969559594989
step: 50, loss: 0.0031166228000074625

```

```

step: 60, loss: 0.1181866005063057
step: 70, loss: 0.00025653530610725284
step: 80, loss: 0.0010115258628502488
step: 90, loss: 0.0005244541680440307
step: 100, loss: 0.0009046756895259023
step: 110, loss: 0.0001265897008124739
step: 120, loss: 0.008692524395883083
step: 130, loss: 0.0036413671914488077
step: 140, loss: 0.01110394299030304
step: 150, loss: 0.0015254253521561623
step: 160, loss: 0.0001768868969520554
step: 170, loss: 0.00013886297529097646
step: 180, loss: 0.0006503167096525431
step: 190, loss: 0.005690506659448147
step: 200, loss: 7.079801434883848e-05
step: 210, loss: 0.00021198992908466607
step: 220, loss: 0.00019521266222000122
step: 230, loss: 0.00017265327915083617
/usr/local/lib/python3.7/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=0.9720044792833147, f1=0.9744160177975528,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.543918763985857e-05
step: 10, loss: 0.0001251143403351307
step: 20, loss: 4.054056989843957e-05
step: 30, loss: 0.0002330330607946962
step: 40, loss: 0.008689201436936855
step: 50, loss: 0.0008312531281262636
step: 60, loss: 0.00016426536603830755
step: 70, loss: 7.820215250831097e-05
step: 80, loss: 0.000584889785386622
step: 90, loss: 0.0007395375869236887
step: 100, loss: 9.616994066163898e-05
step: 110, loss: 0.0008914184290915728
step: 120, loss: 0.00013550787116400898
step: 130, loss: 4.4559659727383405e-05
step: 140, loss: 4.875733066000976e-05
step: 150, loss: 4.462278229766525e-05
step: 160, loss: 0.008960332721471786
step: 170, loss: 0.0053493124432861805
step: 180, loss: 8.695416181581095e-05
step: 190, loss: 0.0016743858577683568
step: 200, loss: 6.838000990683213e-05
step: 210, loss: 5.504085493157618e-05

```



```

step: 220, loss: 5.5853040976217017e-05
step: 230, loss: 0.0012100329622626305
/usr/local/lib/python3.7/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=0.9775784753363228, f1=0.9765363128491621,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023767942911945283
step: 10, loss: 0.0007653216016478837
step: 20, loss: 5.447333387564868e-05
step: 30, loss: 0.0016386094503104687
step: 40, loss: 6.453485548263416e-05
step: 50, loss: 7.047741382848471e-05
step: 60, loss: 0.004878920502960682
step: 70, loss: 6.255399057408795e-05
step: 80, loss: 3.8085145206423476e-05
step: 90, loss: 8.009553130250424e-05
step: 100, loss: 0.0013423701748251915
step: 110, loss: 0.00033749963040463626
step: 120, loss: 0.0003011671360582113
step: 130, loss: 0.0002881035616155714
step: 140, loss: 4.854216967942193e-05
step: 150, loss: 6.367378227878362e-05
step: 160, loss: 0.007460147608071566
step: 170, loss: 9.08052024897188e-05
step: 180, loss: 7.310198270715773e-05
step: 190, loss: 6.216669862624258e-05
step: 200, loss: 0.0055779642425477505
step: 210, loss: 0.00011081140110036358
step: 220, loss: 0.03670985996723175
step: 230, loss: 5.5611191783100367e-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=0.9751693002257337, f1=0.9743016759776536,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.603431782219559e-05
step: 10, loss: 5.315928501659073e-05
step: 20, loss: 9.695332119008526e-05
step: 30, loss: 0.00679824547842145
step: 40, loss: 0.00011614295362960547
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```

```

step: 50, loss: 6.072889664210379e-05
step: 60, loss: 5.027959559811279e-05
step: 70, loss: 0.0001131764511228539
step: 80, loss: 5.449323361972347e-05
step: 90, loss: 3.714734339155257e-05
step: 100, loss: 4.7380039177369326e-05
step: 110, loss: 7.178770465543494e-05
step: 120, loss: 0.0010049488628283143
step: 130, loss: 0.00012715539196506143
step: 140, loss: 0.00662987818941474
step: 150, loss: 0.000326235982356593
step: 160, loss: 4.515184264164418e-05
step: 170, loss: 6.725148705299944e-05
step: 180, loss: 0.00045924054575152695
step: 190, loss: 4.4721287849824876e-05
step: 200, loss: 6.958434823900461e-05
step: 210, loss: 0.0002822648093570024
step: 220, loss: 5.96087375015486e-05
step: 230, loss: 3.19730861519929e-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=0.9766407119021134, f1=0.9723145071982282,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.574997026589699e-05
step: 10, loss: 2.5622013708925806e-05
step: 20, loss: 0.005446490831673145
step: 30, loss: 6.256061169551685e-05
step: 40, loss: 3.665473195724189e-05
step: 50, loss: 0.00010610345634631813
step: 60, loss: 3.7876619899179786e-05
step: 70, loss: 6.095039134379476e-05
step: 80, loss: 7.280670251930133e-05
step: 90, loss: 5.974206578684971e-05
step: 100, loss: 0.004547457210719585
step: 110, loss: 0.00365696894004941
step: 120, loss: 9.932590182870626e-05
step: 130, loss: 8.242343028541654e-05
step: 140, loss: 4.177261143922806e-05
step: 150, loss: 7.21865872037597e-05
step: 160, loss: 3.346286393934861e-05
step: 170, loss: 0.0005262025515548885
step: 180, loss: 8.266337681561708e-05
step: 190, loss: 0.005043736193329096
step: 200, loss: 3.741524778888561e-05

```

```

step: 210, loss: 0.009236360900104046
step: 220, loss: 0.002627963200211525
step: 230, loss: 2.2604714104090817e-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=0.9799107142857142, f1=0.9734513274336283,
best_f1=0.9734513274336283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.061041181557812e-05
step: 10, loss: 7.262515282491222e-05
step: 20, loss: 5.852140748174861e-05
step: 30, loss: 0.0008759470074437559
step: 40, loss: 0.0001604221179150045
step: 50, loss: 8.919773972593248e-05
step: 60, loss: 3.621569703682326e-05
step: 70, loss: 0.00011536006059031934
step: 80, loss: 0.0015479853609576821
step: 90, loss: 4.721105869975872e-05
step: 100, loss: 7.393107080133632e-05
step: 110, loss: 9.498323197476566e-05
step: 120, loss: 5.985624375171028e-05
step: 130, loss: 7.172738696681336e-05
step: 140, loss: 0.012812276370823383
step: 150, loss: 9.099334420170635e-05
step: 160, loss: 0.02007097564637661
step: 170, loss: 3.748575909412466e-05
step: 180, loss: 4.482050280785188e-05
step: 190, loss: 0.0006267100106924772
step: 200, loss: 4.154250200372189e-05
step: 210, loss: 0.001556231640279293
step: 220, loss: 0.0016311134677380323
step: 230, loss: 6.85586201143451e-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=0.9776785714285714, f1=0.9713024282560706,
best_f1=0.9734513274336283
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.")
2473it [00:10, 235.73it/s]
load_f1 = 0.9787709497206705
real_f1 = 0.9753363228699552
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, 254.41it/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: 0.7880210280418396
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4558012783527374
step: 20, loss: 0.49578970670700073
step: 30, loss: 0.4435378909111023
step: 40, loss: 0.31254488229751587
step: 50, loss: 0.21164551377296448
step: 60, loss: 0.14189212024211884
step: 70, loss: 0.07841584086418152
step: 80, loss: 0.12211402505636215
step: 90, loss: 0.17204278707504272
step: 100, loss: 0.2891354262828827
step: 110, loss: 0.0314275398850441
step: 120, loss: 0.0545792356133461
step: 130, loss: 0.020547347143292427
step: 140, loss: 0.03483634442090988
step: 150, loss: 0.02630227990448475
step: 160, loss: 0.13280987739562988
step: 170, loss: 0.1432022899389267
step: 180, loss: 0.07747349888086319
step: 190, loss: 0.020537279546260834
step: 200, loss: 0.08536963909864426
step: 210, loss: 0.06562842428684235
step: 220, loss: 0.08558033406734467
step: 230, loss: 0.0962229073047638
step: 240, loss: 0.05286405235528946
step: 250, loss: 0.06396201252937317
step: 260, loss: 0.07660039514303207
step: 270, loss: 0.013273690827190876
step: 280, loss: 0.06757442653179169
step: 290, loss: 0.01939709112048149
step: 300, loss: 0.29017701745033264
step: 310, loss: 0.05864090099930763
step: 320, loss: 0.03243263438344002
step: 330, loss: 0.04447394981980324
step: 340, loss: 0.07300671935081482
step: 350, loss: 0.07862420380115509

```

```

step: 360, loss: 0.07502826303243637
step: 370, loss: 0.21734532713890076
step: 380, loss: 0.23524920642375946
step: 390, loss: 0.03521803766489029
step: 400, loss: 0.033305492252111435
step: 410, loss: 0.030382351949810982
step: 420, loss: 0.026420805603265762
step: 430, loss: 0.028883084654808044
step: 440, loss: 0.08362017571926117
step: 450, loss: 0.019238866865634918
step: 460, loss: 0.20876353979110718
step: 470, loss: 0.10160231590270996
step: 480, loss: 0.2425450086593628
step: 490, loss: 0.062228988856077194
step: 500, loss: 0.007350745610892773
step: 510, loss: 0.02508445270359516
step: 520, loss: 0.10193168371915817
step: 530, loss: 0.11929730325937271
/usr/local/lib/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=0.9411219286045434, f1=0.9417792268281323,
best_f1=0.9417792268281323
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10969199985265732
step: 10, loss: 0.17052024602890015
step: 20, loss: 0.06517993658781052
step: 30, loss: 0.028690170496702194
step: 40, loss: 0.0034448758233338594
step: 50, loss: 0.028821218758821487
step: 60, loss: 0.07652287930250168
step: 70, loss: 0.17287546396255493
step: 80, loss: 0.007606860250234604
step: 90, loss: 0.0008858045330271125
step: 100, loss: 0.24287569522857666
step: 110, loss: 0.019176362082362175
step: 120, loss: 0.041021913290023804
step: 130, loss: 0.01768181286752224
step: 140, loss: 0.03654072433710098
step: 150, loss: 0.0495486781001091
step: 160, loss: 0.027916766703128815
step: 170, loss: 0.17150358855724335

```

```

step: 180, loss: 0.0030369667802006006
step: 190, loss: 0.02056225761771202
step: 200, loss: 0.007487459108233452
step: 210, loss: 0.004991934634745121
step: 220, loss: 0.0926566943526268
step: 230, loss: 0.016144782304763794
step: 240, loss: 0.06633912771940231
step: 250, loss: 0.01085000578314066
step: 260, loss: 0.05687854811549187
step: 270, loss: 0.07586628198623657
step: 280, loss: 0.09643612802028656
step: 290, loss: 0.11111316084861755
step: 300, loss: 0.058549992740154266
step: 310, loss: 0.0989198163151741
step: 320, loss: 0.015869231894612312
step: 330, loss: 0.014865982346236706
step: 340, loss: 0.004589582793414593
step: 350, loss: 0.029161665588617325
step: 360, loss: 0.003957622218877077
step: 370, loss: 0.006004194263368845
step: 380, loss: 0.03235901892185211
step: 390, loss: 0.0356270857155323
step: 400, loss: 0.010881624184548855
step: 410, loss: 0.0007537746569141746
step: 420, loss: 0.023812711238861084
step: 430, loss: 0.04426830634474754
step: 440, loss: 0.009076067246496677
step: 450, loss: 0.011088748462498188
step: 460, loss: 0.19089852273464203
step: 470, loss: 0.015253094956278801
step: 480, loss: 0.13809160888195038
step: 490, loss: 0.008454689756035805
step: 500, loss: 0.008794363588094711
step: 510, loss: 0.08884605020284653
step: 520, loss: 0.28082242608070374
step: 530, loss: 0.17869850993156433
/usr/local/lib/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=0.945845004668534, f1=0.9416705552963137,
best_f1=0.9416705552963137
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```


step: 0, loss: 0.06238143518567085
step: 10, loss: 0.04730985686182976
step: 20, loss: 0.07581702619791031
step: 30, loss: 0.17240585386753082
step: 40, loss: 0.02152123674750328
step: 50, loss: 0.0036820792593061924
step: 60, loss: 0.0011743120849132538
step: 70, loss: 0.002592227188870311
step: 80, loss: 0.0010469313710927963
step: 90, loss: 0.019648397341370583
step: 100, loss: 0.1710471659898758
step: 110, loss: 0.007012067828327417
step: 120, loss: 0.03936845436692238
step: 130, loss: 0.021071458235383034
step: 140, loss: 0.03536571189761162
step: 150, loss: 0.007407679222524166
step: 160, loss: 0.008186164312064648
step: 170, loss: 0.0467422753572464
step: 180, loss: 0.014975124038755894
step: 190, loss: 0.03897016495466232
step: 200, loss: 0.0020741529297083616
step: 210, loss: 0.008948013186454773
step: 220, loss: 0.037671640515327454
step: 230, loss: 0.05372196063399315
step: 240, loss: 0.05184078961610794
step: 250, loss: 0.0021123094484210014
step: 260, loss: 0.0014619934372603893
step: 270, loss: 0.0018245266983285546
step: 280, loss: 0.005776519421488047
step: 290, loss: 0.08259252458810806
step: 300, loss: 0.11891961842775345
step: 310, loss: 0.09516286849975586
step: 320, loss: 0.1173921674489975
step: 330, loss: 0.00529109500348568
step: 340, loss: 0.006353450007736683
step: 350, loss: 0.009451664052903652
step: 360, loss: 0.010031195357441902
step: 370, loss: 0.0010538523783907294
step: 380, loss: 0.0038824756629765034
step: 390, loss: 0.052027370780706406
step: 400, loss: 0.0396636538207531
step: 410, loss: 0.01344163529574871
step: 420, loss: 0.09165345132350922
step: 430, loss: 0.014731678180396557
step: 440, loss: 0.015016397461295128
step: 450, loss: 0.05714261904358864
step: 460, loss: 0.05221542343497276
step: 470, loss: 0.0016685118898749352

```

step: 480, loss: 0.011123519390821457
step: 490, loss: 0.010640686377882957
step: 500, loss: 0.06833001971244812
step: 510, loss: 0.0013104280224069953
step: 520, loss: 0.003989624325186014
step: 530, loss: 0.09311831742525101
/usr/local/lib/python3.7/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=0.944954128440367, f1=0.939435968562182,
best_f1=0.9416705552963137
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011485809227451682
step: 10, loss: 0.008910593576729298
step: 20, loss: 0.010564486496150494
step: 30, loss: 0.004405755549669266
step: 40, loss: 0.03915497288107872
step: 50, loss: 0.07146157324314117
step: 60, loss: 0.0008631751406937838
step: 70, loss: 0.016086390241980553
step: 80, loss: 0.04890302196145058
step: 90, loss: 0.0047990973107516766
step: 100, loss: 0.007880917750298977
step: 110, loss: 0.0010951516451314092
step: 120, loss: 0.0009039563592523336
step: 130, loss: 0.01962534338235855
step: 140, loss: 0.019025586545467377
step: 150, loss: 0.0007547244313172996
step: 160, loss: 0.0005635849665850401
step: 170, loss: 0.06382995843887329
step: 180, loss: 0.001298930961638689
step: 190, loss: 0.005197578575462103
step: 200, loss: 0.000909114838577807
step: 210, loss: 0.13840703666210175
step: 220, loss: 0.0028564154636114836
step: 230, loss: 0.3101971745491028
step: 240, loss: 0.003130018012598157
step: 250, loss: 0.006870690733194351
step: 260, loss: 0.10948450118303299
step: 270, loss: 0.0026342077180743217
step: 280, loss: 0.004323300905525684
step: 290, loss: 0.0023587760515511036
step: 300, loss: 0.0022110084537416697
step: 310, loss: 0.0016368450596928596
step: 320, loss: 0.019584139809012413
step: 330, loss: 0.01711081899702549

```

```

step: 340, loss: 0.0009544738568365574
step: 350, loss: 0.00044207615428604186
step: 360, loss: 0.0029857621993869543
step: 370, loss: 0.011358690448105335
step: 380, loss: 0.0005957218236289918
step: 390, loss: 0.009375563822686672
step: 400, loss: 0.0053574093617498875
step: 410, loss: 0.002639233833178878
step: 420, loss: 0.006620639469474554
step: 430, loss: 0.02042868733406067
step: 440, loss: 0.00768235744908452
step: 450, loss: 0.07985125482082367
step: 460, loss: 0.0030894302763044834
step: 470, loss: 0.002116747433319688
step: 480, loss: 0.06055315211415291
step: 490, loss: 0.06325297802686691
step: 500, loss: 0.0035988136660307646
step: 510, loss: 0.019363321363925934
step: 520, loss: 0.010806330479681492
step: 530, loss: 0.0011712451232597232
/usr/local/lib/python3.7/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=0.9441624365482235, f1=0.9422632794457274,
best_f1=0.9416705552963137
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028854981064796448
step: 10, loss: 0.01971842162311077
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 0.014253824017941952
step: 30, loss: 0.014485981315374374
step: 40, loss: 0.05519455671310425
step: 50, loss: 0.002355834236368537
step: 60, loss: 0.0011683186748996377
step: 70, loss: 0.0018444785382598639
step: 80, loss: 0.0012535849818959832
step: 90, loss: 0.019215697422623634
step: 100, loss: 0.0003668577701319009
step: 110, loss: 0.0017279890598729253
step: 120, loss: 0.002104339189827442
step: 130, loss: 0.0004054927558172494
step: 140, loss: 0.0005328806000761688
step: 150, loss: 0.000589993316680193
step: 160, loss: 0.0022494818549603224
step: 170, loss: 0.04010140150785446
step: 180, loss: 0.0035795154981315136

```

```

step: 190, loss: 0.000126565239042975
step: 200, loss: 0.0006250359583646059
step: 210, loss: 0.0016143731772899628
step: 220, loss: 0.0002950254420284182
step: 230, loss: 0.0027374245692044497
step: 240, loss: 0.0009947835933417082
step: 250, loss: 0.0008658197475597262
step: 260, loss: 0.08047743886709213
step: 270, loss: 0.0006811016937717795
step: 280, loss: 0.08436061441898346
step: 290, loss: 0.0010321909794583917
step: 300, loss: 0.10115543007850647
step: 310, loss: 0.0010228934697806835
step: 320, loss: 0.0005593172390945256
step: 330, loss: 0.00013691659842152148
step: 340, loss: 0.0030272623989731073
step: 350, loss: 0.00024631727137602866
step: 360, loss: 0.003921207040548325
step: 370, loss: 0.0011230645468458533
step: 380, loss: 0.06677044928073883
step: 390, loss: 0.01809459738433361
step: 400, loss: 0.0032227966003119946
step: 410, loss: 0.0008729806286282837
step: 420, loss: 0.0014395862817764282
step: 430, loss: 0.0010586556745693088
step: 440, loss: 0.0015761306276544929
step: 450, loss: 0.00987178273499012
step: 460, loss: 0.003920596092939377
step: 470, loss: 0.007757342420518398
step: 480, loss: 0.0027878510300070047
step: 490, loss: 0.06280151009559631
step: 500, loss: 0.0317157618701458
step: 510, loss: 0.0015286959242075682
step: 520, loss: 0.0002740098279900849
step: 530, loss: 0.021902140229940414
/usr/local/lib/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=0.9482439926062847, f1=0.9445983379501385,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003204751992598176

```

step: 10, loss: 0.005350218154489994
step: 20, loss: 0.0013362632598727942
step: 30, loss: 0.011900500394403934
step: 40, loss: 0.0009302037069573998
step: 50, loss: 0.043121661990880966
step: 60, loss: 0.0006789744948036969
step: 70, loss: 0.001481826649978757
step: 80, loss: 0.0007258846308104694
step: 90, loss: 0.00044515274930745363
step: 100, loss: 0.0331764779984951
step: 110, loss: 0.0016039354959502816
step: 120, loss: 0.0008867710130289197
step: 130, loss: 0.00024614352150820196
step: 140, loss: 0.09486132860183716
step: 150, loss: 0.004571182187646627
step: 160, loss: 0.0007707307231612504
step: 170, loss: 9.233332093572244e-05
step: 180, loss: 6.0651247622445226e-05
step: 190, loss: 0.0008171928348019719
step: 200, loss: 0.0043242997489869595
step: 210, loss: 0.00019420104217715561
step: 220, loss: 0.001492758747190237
step: 230, loss: 0.0002885431458707899
step: 240, loss: 0.02267630212008953
step: 250, loss: 0.0003517967415973544
step: 260, loss: 0.0003427816554903984
step: 270, loss: 0.019410917535424232
step: 280, loss: 0.003145997179672122
step: 290, loss: 0.07112520188093185
step: 300, loss: 0.00015707073907833546
step: 310, loss: 0.00023483902623411268
step: 320, loss: 0.0036464680451899767
step: 330, loss: 0.008554334752261639
step: 340, loss: 0.0005236385040916502
step: 350, loss: 0.027170775458216667
step: 360, loss: 0.0016222040867432952
step: 370, loss: 0.003303012577816844
step: 380, loss: 0.11064959317445755
step: 390, loss: 0.11259083449840546
step: 400, loss: 0.0017140187555924058
step: 410, loss: 0.00032922718673944473
step: 420, loss: 0.00443640723824501
step: 430, loss: 0.0016511034918949008
step: 440, loss: 0.0020431240554898977
step: 450, loss: 0.0013884976506233215
step: 460, loss: 0.0017197778215631843
step: 470, loss: 0.006275043822824955
step: 480, loss: 0.02259472757577896

```

step: 490, loss: 0.00017806206597015262
step: 500, loss: 0.0309067964553833
step: 510, loss: 0.0002124305028701201
step: 520, loss: 0.13670207560062408
step: 530, loss: 0.0013211783953011036
/usr/local/lib/python3.7/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=0.9387755102040817, f1=0.9429234338747099,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002920735627412796
step: 10, loss: 0.008162728510797024
step: 20, loss: 0.12135009467601776
step: 30, loss: 0.010100819170475006
step: 40, loss: 6.714123446727172e-05
step: 50, loss: 0.0009034068207256496
step: 60, loss: 0.011716215871274471
step: 70, loss: 0.07681621611118317
step: 80, loss: 0.01579578034579754
step: 90, loss: 0.0005324315279722214
step: 100, loss: 0.0005271168192848563
step: 110, loss: 0.0022915725130587816
step: 120, loss: 0.00014545049634762108
step: 130, loss: 0.01385459303855896
step: 140, loss: 0.005705852527171373
step: 150, loss: 7.687724428251386e-05
step: 160, loss: 0.0008160074939951301
step: 170, loss: 0.0008240028982982039
step: 180, loss: 0.0003546452207956463
step: 190, loss: 8.849799633026123e-05
step: 200, loss: 0.0002983784943353385
step: 210, loss: 0.00014549923071172088
step: 220, loss: 0.0006516727153211832
step: 230, loss: 0.00032036102493293583
step: 240, loss: 0.0015238674823194742
step: 250, loss: 0.004351533018052578
step: 260, loss: 0.0009372580680064857
step: 270, loss: 0.00048048445023596287
step: 280, loss: 0.005239068530499935
step: 290, loss: 0.0029303624760359526
step: 300, loss: 0.0009426686447113752
step: 310, loss: 0.001116391969844699
step: 320, loss: 0.011380312032997608
step: 330, loss: 0.0010448978282511234
step: 340, loss: 0.13121990859508514

```

```

step: 350, loss: 0.0232082586735487
step: 360, loss: 0.0032702197786420584
step: 370, loss: 0.0018297320930287242
step: 380, loss: 0.006567784119397402
step: 390, loss: 8.102035644697025e-05
step: 400, loss: 0.00010125218977918848
step: 410, loss: 0.0014424598775804043
step: 420, loss: 0.0006631847354583442
step: 430, loss: 7.096683839336038e-05
step: 440, loss: 7.671565253986046e-05
step: 450, loss: 0.0090890908613801
step: 460, loss: 0.0011827574344351888
step: 470, loss: 0.004599595442414284
step: 480, loss: 0.00014873383042868227
step: 490, loss: 0.000698078132700175
step: 500, loss: 0.0038095437921583652
step: 510, loss: 0.0009594660950824618
step: 520, loss: 0.0002178275171900168
step: 530, loss: 9.166301606455818e-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=0.9433962264150944, f1=0.9384835479256081,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009243926033377647
step: 10, loss: 0.0032604869920760393
step: 20, loss: 0.012521347030997276
step: 30, loss: 0.00014616927364841104
step: 40, loss: 0.0010419196914881468
step: 50, loss: 0.000203883319045417
step: 60, loss: 4.748280480271205e-05
step: 70, loss: 0.04898783192038536
step: 80, loss: 0.000486047996673733
step: 90, loss: 0.003822661703452468
step: 100, loss: 0.0009489086223766208
step: 110, loss: 0.0009104878408834338
step: 120, loss: 0.00014730944531038404
step: 130, loss: 0.1964220106601715
step: 140, loss: 2.9674400138901547e-05
step: 150, loss: 0.012508932501077652
step: 160, loss: 0.00032840241328813136
step: 170, loss: 0.0004338521684985608
step: 180, loss: 0.00021703731908928603
step: 190, loss: 0.00039304789970628917
step: 200, loss: 0.00013770964869763702

```

```

step: 210, loss: 0.0008167822961695492
step: 220, loss: 0.001160759013146162
step: 230, loss: 0.00018316417117603123
step: 240, loss: 0.00034283558488823473
step: 250, loss: 0.000307970040012151
step: 260, loss: 0.0013781816232949495
step: 270, loss: 3.955170541303232e-05
step: 280, loss: 0.0030526542104780674
step: 290, loss: 0.0001823667698772624
step: 300, loss: 0.00012613651051651686
step: 310, loss: 0.013640111312270164
step: 320, loss: 0.004895028658211231
step: 330, loss: 0.04543082416057587
step: 340, loss: 0.0030086576007306576
step: 350, loss: 0.12490388005971909
step: 360, loss: 0.00020265883358661085
step: 370, loss: 0.0003851034271065146
step: 380, loss: 0.002175732050091028
step: 390, loss: 0.000738775182981044
step: 400, loss: 0.12313626706600189
step: 410, loss: 0.0005298135220073164
step: 420, loss: 0.0016175543423742056
step: 430, loss: 0.0002913489588536322
step: 440, loss: 0.0021185989025980234
step: 450, loss: 0.0020628126803785563
step: 460, loss: 0.0021273770835250616
step: 470, loss: 0.0013362063327804208
step: 480, loss: 0.0009202480432577431
step: 490, loss: 0.0004944750689901412
step: 500, loss: 0.004550849553197622
step: 510, loss: 0.0002639327140059322
step: 520, loss: 0.00012024449097225443
step: 530, loss: 0.0005996161489747465
/usr/local/lib/python3.7/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=0.9442139234670355, f1=0.9425287356321841,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002529521007090807
step: 10, loss: 0.0011716028675436974
step: 20, loss: 0.00037747336318716407
step: 30, loss: 0.004019766114652157
step: 40, loss: 0.00020364069496281445
step: 50, loss: 0.00012289013830013573
step: 60, loss: 0.0002310755371581763

```


step: 70, loss: 0.12080906331539154
step: 80, loss: 0.0003418239939492196
step: 90, loss: 0.0015975292772054672
step: 100, loss: 0.0011125619057565928
step: 110, loss: 0.03462131321430206
step: 120, loss: 5.2886945923091844e-05
step: 130, loss: 6.559486064361408e-05
step: 140, loss: 0.00017095702060032636
step: 150, loss: 9.951261745300144e-05
step: 160, loss: 7.106201519491151e-05
step: 170, loss: 4.0450864617014304e-05
step: 180, loss: 0.00020105561998207122
step: 190, loss: 0.00011829522554762661
step: 200, loss: 0.00020072460756637156
step: 210, loss: 7.726255716988817e-05
step: 220, loss: 3.848047708743252e-05
step: 230, loss: 0.0025651706382632256
step: 240, loss: 6.89813241478987e-05
step: 250, loss: 0.00028568465495482087
step: 260, loss: 0.00018545197963248938
step: 270, loss: 0.007227376103401184
step: 280, loss: 4.4868043914902955e-05
step: 290, loss: 3.407974145375192e-05
step: 300, loss: 0.00010177902731811628
step: 310, loss: 0.00015212058497127146
step: 320, loss: 0.0005571643123403192
step: 330, loss: 0.02185834012925625
step: 340, loss: 3.367730460013263e-05
step: 350, loss: 4.2282834328943864e-05
step: 360, loss: 0.04592077434062958
step: 370, loss: 0.00014817703049629927
step: 380, loss: 0.00012851040810346603
step: 390, loss: 2.8195558115839958e-05
step: 400, loss: 0.06176459789276123
step: 410, loss: 0.0011681296164169908
step: 420, loss: 0.002190865809097886
step: 430, loss: 3.8020902138669044e-05
step: 440, loss: 5.032201806898229e-05
step: 450, loss: 8.521180279785767e-05
step: 460, loss: 0.0009055117843672633
step: 470, loss: 3.266183193773031e-05
step: 480, loss: 0.0002790557045955211
step: 490, loss: 0.0003060148737858981
step: 500, loss: 0.00014761753845959902
step: 510, loss: 0.00021074323740322143
step: 520, loss: 9.741303802002221e-05
step: 530, loss: 0.00012945078196935356
/usr/local/lib/python3.7/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=0.9480519480519481, f1=0.9459962756052142,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007736764382570982
step: 10, loss: 2.3002206944511272e-05
step: 20, loss: 0.0001249659835593775
step: 30, loss: 9.394515655003488e-05
step: 40, loss: 8.076818630797789e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 5.495197183336131e-05
step: 60, loss: 6.641587970079854e-05
step: 70, loss: 0.0003343858988955617
step: 80, loss: 0.0011623569298535585
step: 90, loss: 0.00034659591619856656
step: 100, loss: 0.0003411528596188873
step: 110, loss: 0.0009111157851293683
step: 120, loss: 4.189580795355141e-05
step: 130, loss: 2.6199122657999396e-05
step: 140, loss: 7.654061482753605e-05
step: 150, loss: 4.642034036805853e-05
step: 160, loss: 0.00013485908857546747
step: 170, loss: 5.487330781761557e-05
step: 180, loss: 0.0002991882502101362
step: 190, loss: 0.00013866613153368235
step: 200, loss: 6.998553726589307e-05
step: 210, loss: 7.672778883716092e-05
step: 220, loss: 3.6498568078968674e-05
step: 230, loss: 0.00121508096344769
step: 240, loss: 0.0001462022337364033
step: 250, loss: 0.00021385734726209193
step: 260, loss: 0.00203879876062274
step: 270, loss: 0.004027035087347031
step: 280, loss: 0.00013140079681761563
step: 290, loss: 7.547948916908354e-05
step: 300, loss: 0.011805538088083267
step: 310, loss: 0.0003402756992727518
step: 320, loss: 7.003324571996927e-05
step: 330, loss: 4.2748441046569496e-05
step: 340, loss: 3.345442019053735e-05
step: 350, loss: 8.64961912157014e-05
step: 360, loss: 0.0007377448491752148
step: 370, loss: 0.0002506535092834383
step: 380, loss: 0.000141965807415545
step: 390, loss: 6.992428825469688e-05
```

```

step: 400, loss: 0.00021269070566631854
step: 410, loss: 0.00048125398461706936
step: 420, loss: 0.00019971404981333762
step: 430, loss: 0.00010325961920898408
step: 440, loss: 5.144619717611931e-05
step: 450, loss: 0.006262808572500944
step: 460, loss: 0.0033250530250370502
step: 470, loss: 2.6921627068077214e-05
step: 480, loss: 0.04912136495113373
step: 490, loss: 0.02434265799820423
step: 500, loss: 0.0020418378990143538
step: 510, loss: 0.0001609546598047018
step: 520, loss: 2.7655392841552384e-05
step: 530, loss: 5.2071856771362945e-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=0.946146703806871, f1=0.9475638051044084,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026051552966237068
step: 10, loss: 0.0010331544326618314
step: 20, loss: 0.000491177779622376
step: 30, loss: 0.00194224261213094
step: 40, loss: 0.00011471026664366946
step: 50, loss: 9.42840488278307e-05
step: 60, loss: 0.00013198252418078482
step: 70, loss: 5.272932321531698e-05
step: 80, loss: 0.0014001333620399237
step: 90, loss: 0.0040205661207437515
step: 100, loss: 0.0006789295584894717
step: 110, loss: 0.07483454048633575
step: 120, loss: 0.00016290144412778318
step: 130, loss: 3.0326067644637078e-05
step: 140, loss: 0.00010261899296892807
step: 150, loss: 8.285331568913534e-05
step: 160, loss: 0.00020861939992755651
step: 170, loss: 0.00032785782241262496
step: 180, loss: 3.632668813224882e-05
step: 190, loss: 0.00022514727606903762
step: 200, loss: 0.0029345317743718624
step: 210, loss: 5.1955285016447306e-05
step: 220, loss: 4.095723124919459e-05
step: 230, loss: 0.0003252651949878782
step: 240, loss: 0.00031022983603179455
step: 250, loss: 5.097558823763393e-05

```

```

step: 260, loss: 2.7949279683525674e-05
step: 270, loss: 0.0009624222875572741
step: 280, loss: 6.619955820497125e-05
step: 290, loss: 0.0024000119883567095
step: 300, loss: 0.00048495197552256286
step: 310, loss: 0.00022381196322385222
step: 320, loss: 0.03635309636592865
step: 330, loss: 0.0008114826632663608
step: 340, loss: 0.13706475496292114
step: 350, loss: 0.00047484852257184684
step: 360, loss: 0.00018596203881315887
step: 370, loss: 3.2948071748251095e-05
step: 380, loss: 5.0142494728788733e-05
step: 390, loss: 0.0035372753627598286
step: 400, loss: 5.377620254876092e-05
step: 410, loss: 6.962223415030167e-05
step: 420, loss: 8.375387551495805e-05
step: 430, loss: 0.001611936604604125
step: 440, loss: 6.103995838202536e-05
step: 450, loss: 7.944936805870384e-05
step: 460, loss: 0.001280666678212583
step: 470, loss: 0.03434516489505768
step: 480, loss: 7.066542457323521e-05
step: 490, loss: 0.0014354379381984472
step: 500, loss: 5.3674812079407275e-05
step: 510, loss: 4.597449515131302e-05
step: 520, loss: 4.6406341425608844e-05
step: 530, loss: 2.498485628166236e-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=0.943239501615136, f1=0.9405803777061263,
best_f1=0.9445983379501385
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008396786288358271
step: 10, loss: 0.0007199224201031029
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 20, loss: 2.829943878168706e-05
step: 30, loss: 3.644329262897372e-05
step: 40, loss: 0.00013723305892199278
step: 50, loss: 0.0007489334675483406
step: 60, loss: 0.0028021549805998802
step: 70, loss: 0.0001377102016704157
step: 80, loss: 0.0014178149867802858
step: 90, loss: 0.00011376449401723221
step: 100, loss: 0.00014660466695204377

```

```
step: 110, loss: 2.2764108507544734e-05
step: 120, loss: 0.00012265692930668592
step: 130, loss: 5.647173020406626e-05
step: 140, loss: 6.108533852966502e-05
step: 150, loss: 0.0006859907298348844
step: 160, loss: 0.00011910560715477914
step: 170, loss: 6.49418871034868e-05
step: 180, loss: 0.00018824446306098253
step: 190, loss: 0.00018340827955398709
step: 200, loss: 0.0002537354885134846
step: 210, loss: 0.0005122975562699139
step: 220, loss: 4.754212568514049e-05
step: 230, loss: 0.0013779210858047009
step: 240, loss: 7.651546911802143e-05
step: 250, loss: 0.00017421224038116634
step: 260, loss: 9.499488078290597e-05
step: 270, loss: 0.0018714000470936298
step: 280, loss: 2.8512076823972166e-05
step: 290, loss: 0.012154903262853622
step: 300, loss: 0.00010711923096096143
step: 310, loss: 3.952326369471848e-05
step: 320, loss: 4.898674524156377e-05
step: 330, loss: 4.713940870715305e-05
step: 340, loss: 7.504018867621198e-05
step: 350, loss: 0.0038155447691679
step: 360, loss: 0.00023937440710142255
step: 370, loss: 2.296576894877944e-05
step: 380, loss: 8.203570905607194e-05
step: 390, loss: 3.240010482841171e-05
step: 400, loss: 6.514137203339487e-05
step: 410, loss: 0.0002516317181289196
step: 420, loss: 0.0007590859313495457
step: 430, loss: 9.817094542086124e-05
step: 440, loss: 0.03268450126051903
step: 450, loss: 0.0035380437038838863
step: 460, loss: 6.718836084473878e-05
step: 470, loss: 0.0001249877386726439
step: 480, loss: 0.00013159375521354377
step: 490, loss: 4.054418604937382e-05
step: 500, loss: 0.0013863538624718785
step: 510, loss: 0.0003963317722082138
step: 520, loss: 0.00965385977178812
step: 530, loss: 0.0005344062228687108
/usr/local/lib/python3.7/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=0.9467289719626168, f1=0.9444184960298926,
best_f1=0.9445983379501385
```

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

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

step: 0, loss: 0.000543645117431879
step: 10, loss: 0.00029162393184378743
step: 20, loss: 0.004680966027081013
step: 30, loss: 0.00023335727746598423
step: 40, loss: 3.637883492046967e-05
step: 50, loss: 0.00015067856293171644
step: 60, loss: 0.0034175203181803226
step: 70, loss: 3.8501122617162764e-05
step: 80, loss: 2.3699449229752645e-05
step: 90, loss: 0.0012175623560324311
step: 100, loss: 2.0603933080565184e-05
step: 110, loss: 6.93581678206101e-05
step: 120, loss: 4.415232251631096e-05
step: 130, loss: 0.001724116737022996
step: 140, loss: 7.803007611073554e-05
step: 150, loss: 6.293976912274957e-05
step: 160, loss: 2.279814907524269e-05
step: 170, loss: 7.798031583661214e-05
step: 180, loss: 2.4682763978489675e-05
step: 190, loss: 0.002208529505878687
step: 200, loss: 0.002773007145151496
step: 210, loss: 0.0008741223718971014
step: 220, loss: 3.482034298940562e-05
step: 230, loss: 0.00014216946146916598
step: 240, loss: 5.5233649618458e-05
step: 250, loss: 0.03279915079474449
step: 260, loss: 2.967773616546765e-05
step: 270, loss: 3.6904544685967267e-05
step: 280, loss: 1.4088840543990955e-05
step: 290, loss: 4.0919323510024697e-05
step: 300, loss: 4.823434574063867e-05
step: 310, loss: 0.006136063951998949
step: 320, loss: 5.412744212662801e-05
step: 330, loss: 0.0006640666979365051
step: 340, loss: 2.3874828912084922e-05
step: 350, loss: 0.008637289516627789
step: 360, loss: 2.299170955666341e-05
step: 370, loss: 1.9128921849187464e-05
step: 380, loss: 0.002751752967014909
step: 390, loss: 0.0012163093779236078
step: 400, loss: 2.7990094167762436e-05
step: 410, loss: 3.314171408419497e-05
step: 420, loss: 4.851278572459705e-05
step: 430, loss: 0.0001659196859691292
step: 440, loss: 0.0004162554978393018

```

step: 450, loss: 0.0019994997419416904
step: 460, loss: 3.561206904123537e-05
step: 470, loss: 0.00017346769163850695
step: 480, loss: 0.0007117632776498795
step: 490, loss: 2.6239864382660016e-05
step: 500, loss: 0.000137133858515881
step: 510, loss: 4.554527913569473e-05
step: 520, loss: 2.8210160962771624e-05
step: 530, loss: 2.9253362299641594e-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 13: dev_f1=0.9485396383866481, f1=0.9478060046189377,
best_f1=0.9478060046189377
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.055041790707037e-05
step: 10, loss: 3.056083369301632e-05
step: 20, loss: 1.783262996468693e-05
step: 30, loss: 3.280733653809875e-05
step: 40, loss: 0.0003441088192630559
step: 50, loss: 6.788618338759989e-05
step: 60, loss: 3.231531445635483e-05
step: 70, loss: 0.00043318033567629755
step: 80, loss: 7.81975977588445e-05
step: 90, loss: 2.197793946834281e-05
step: 100, loss: 5.7982739235740155e-05
step: 110, loss: 2.4954726541182026e-05
step: 120, loss: 1.948264252860099e-05
step: 130, loss: 4.183947021374479e-05
step: 140, loss: 0.0024268697015941143
step: 150, loss: 5.560059435083531e-05
step: 160, loss: 0.0032094777561724186
step: 170, loss: 0.00014910916797816753
step: 180, loss: 0.0011723325587809086
step: 190, loss: 0.00040925171924754977
step: 200, loss: 3.966252552345395e-05
step: 210, loss: 0.00022502704814542085
step: 220, loss: 2.8652846594923176e-05
step: 230, loss: 0.0009298124350607395
step: 240, loss: 8.232032996602356e-05
step: 250, loss: 4.777380672749132e-05
step: 260, loss: 1.4073969396122266e-05

```

```

step: 270, loss: 0.03658873960375786
step: 280, loss: 0.00020035143825225532
step: 290, loss: 0.0005380940274335444
step: 300, loss: 0.0001113399412133731
step: 310, loss: 2.860749555111397e-05
step: 320, loss: 2.6970192266162485e-05
step: 330, loss: 4.07749685109593e-05
step: 340, loss: 3.714577906066552e-05
step: 350, loss: 0.0004177349910605699
step: 360, loss: 0.0016259270487353206
step: 370, loss: 2.8273052521399222e-05
step: 380, loss: 2.1996716895955615e-05
step: 390, loss: 0.00011690715473378077
step: 400, loss: 3.26021654473152e-05
step: 410, loss: 2.3155695089371875e-05
step: 420, loss: 3.568617466953583e-05
step: 430, loss: 2.631782081152778e-05
step: 440, loss: 1.4655109225714114e-05
step: 450, loss: 2.0306171791162342e-05
step: 460, loss: 4.988432920072228e-05
step: 470, loss: 3.742015906027518e-05
step: 480, loss: 2.9812013963237405e-05
step: 490, loss: 1.994844751607161e-05
step: 500, loss: 4.686527972808108e-05
step: 510, loss: 3.06496222037822e-05
step: 520, loss: 7.727097545284778e-05
step: 530, loss: 2.1062403902760707e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 4096.0
/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 14: dev_f1=0.9456928838951311, f1=0.9451476793248945,
best_f1=0.9478060046189377
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.5014218408614397e-05
step: 10, loss: 6.447068153647706e-05
step: 20, loss: 8.345690002897754e-05
step: 30, loss: 0.0005298594478517771
step: 40, loss: 1.245723160536727e-05
step: 50, loss: 0.00016466517990920693
step: 60, loss: 3.4229986340506e-05
step: 70, loss: 6.43419407424517e-05
step: 80, loss: 1.794811396393925e-05
step: 90, loss: 3.572465720935725e-05
step: 100, loss: 6.315656355582178e-05
step: 110, loss: 0.019920766353607178

```



```

step: 120, loss: 0.001992368372157216
step: 130, loss: 1.435332524124533e-05
step: 140, loss: 1.6815522030810826e-05
step: 150, loss: 0.0011342610232532024
step: 160, loss: 0.00048822929966263473
step: 170, loss: 3.515884236549027e-05
step: 180, loss: 2.734910049184691e-05
step: 190, loss: 9.19573794817552e-05
step: 200, loss: 3.2555464713368565e-05
step: 210, loss: 1.8175118384533562e-05
step: 220, loss: 1.7296046280534938e-05
step: 230, loss: 2.267871968797408e-05
step: 240, loss: 4.425721272127703e-05
step: 250, loss: 0.11088885366916656
step: 260, loss: 0.00011927072773687541
step: 270, loss: 0.0002460437244735658
step: 280, loss: 3.95108581869863e-05
step: 290, loss: 0.0007653250941075385
step: 300, loss: 0.0038102450780570507
step: 310, loss: 9.5345065346919e-05
step: 320, loss: 5.9821639297297224e-05
step: 330, loss: 4.0025483031058684e-05
step: 340, loss: 0.02381422184407711
step: 350, loss: 5.622317985398695e-05
step: 360, loss: 0.005256905686110258
step: 370, loss: 2.228390621894505e-05
step: 380, loss: 1.773206349753309e-05
step: 390, loss: 2.328180562471971e-05
step: 400, loss: 1.560872988193296e-05
step: 410, loss: 0.0008745858212932944
step: 420, loss: 8.439324301434681e-05
step: 430, loss: 2.3885864720796235e-05
step: 440, loss: 0.003624033648520708
step: 450, loss: 0.0023319164756685495
step: 460, loss: 1.5318166333599947e-05
step: 470, loss: 1.8663235096028075e-05
step: 480, loss: 0.00035042857052758336
step: 490, loss: 4.3928885133937e-05
step: 500, loss: 0.00015228893607854843
step: 510, loss: 0.0005821814411319792
step: 520, loss: 7.111232116585597e-05
step: 530, loss: 6.423239392461255e-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=0.9451360073766714, f1=0.942095588235294,
best_f1=0.9478060046189377
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, 254.82it/s]

load_f1 = 0.9499536607970344

real_f1 = 0.9480037140204272

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, 254.82it/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: 0.8634424209594727
/usr/local/lib/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=0.35443037974683544, f1=0.35443037974683544,
best_f1=0.35443037974683544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.36258694529533386
/usr/local/lib/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=0.4210526315789474, f1=0.3414634146341463,
best_f1=0.3414634146341463
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32603713870048523
/usr/local/lib/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=0.4444444444444444, f1=0.35000000000000003,
best_f1=0.35000000000000003
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.33479413390159607
/usr/local/lib/python3.7/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=0.41935483870967744, f1=0.4, best_f1=0.35000000000000003
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25065305829048157
/usr/local/lib/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=0.4727272727272727, f1=0.42857142857142855,
best_f1=0.42857142857142855
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.246168315410614
/usr/local/lib/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=0.4905660377358491, f1=0.43137254901960786,
best_f1=0.43137254901960786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2997584939002991
/usr/local/lib/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=0.5714285714285714, f1=0.5, best_f1=0.5
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.49887895584106445
/usr/local/lib/python3.7/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=0.5652173913043478, f1=0.43999999999999995, best_f1=0.5
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17258483171463013
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 9: dev_f1=0.6285714285714286, f1=0.4864864864864865,
best_f1=0.4864864864864865
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22791309654712677
/usr/local/lib/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=0.7096774193548386, f1=0.5142857142857143,
best_f1=0.5142857142857143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20469582080841064
/usr/local/lib/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=0.7857142857142857, f1=0.5454545454545454,
best_f1=0.5454545454545454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24245786666870117
/usr/local/lib/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=0.7999999999999999, f1=0.5714285714285714,
best_f1=0.5714285714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13708755373954773
/usr/local/lib/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=0.8275862068965518, f1=0.5555555555555556,
best_f1=0.5555555555555556
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15400779247283936
/usr/local/lib/python3.7/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=0.8275862068965518, f1=0.5714285714285714,
best_f1=0.5555555555555556
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24552972614765167
/usr/local/lib/python3.7/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=0.8275862068965518, f1=0.5714285714285714,
best_f1=0.5555555555555556
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, 129252.17it/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 = 0.7333333333333334
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, 239.70it/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/
```

```
Downloading: 100% 442/442 [00:00<00:00, 397kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 349kB/s]
Downloading: 100% 268M/268M [00:04<00:00, 58.5MB/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: 0.8063173890113831
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4586564600467682
step: 20, loss: 0.605564534664154
step: 30, loss: 0.4299333691596985
step: 40, loss: 0.20745520293712616
step: 50, loss: 0.0488537959754467
step: 60, loss: 0.0871925875544548
step: 70, loss: 0.10976861417293549
step: 80, loss: 0.21155433356761932
step: 90, loss: 0.009970569051802158
step: 100, loss: 0.05446609482169151
step: 110, loss: 0.045167192816734314
step: 120, loss: 0.020998047664761543
step: 130, loss: 0.004318676423281431
step: 140, loss: 0.010387502610683441
step: 150, loss: 0.09088078886270523
step: 160, loss: 0.1114625632762909
step: 170, loss: 0.007182457949966192
step: 180, loss: 0.00858269166201353

```

```

step: 190, loss: 0.32468393445014954
step: 200, loss: 0.011268987320363522
step: 210, loss: 0.009920196607708931
step: 220, loss: 0.005707936827093363
step: 230, loss: 0.0043088593520224094
/usr/local/lib/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=0.984304932735426, f1=0.9841986455981941,
best_f1=0.9841986455981941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004063945263624191
step: 10, loss: 0.0012309147277846932
step: 20, loss: 0.0010935781756415963
step: 30, loss: 0.022142842411994934
step: 40, loss: 0.004711305722594261
step: 50, loss: 0.005370438564568758
step: 60, loss: 0.002731951652094722
step: 70, loss: 0.005268091335892677
step: 80, loss: 0.0012718824436888099
step: 90, loss: 0.007123507093638182
step: 100, loss: 0.06873670965433121
step: 110, loss: 0.03225003927946091
step: 120, loss: 0.0016143822576850653
step: 130, loss: 0.0015667567495256662
step: 140, loss: 0.30105075240135193
step: 150, loss: 0.006334810983389616
step: 160, loss: 0.003071045270189643
step: 170, loss: 0.00621007988229394
step: 180, loss: 0.004090556409209967
step: 190, loss: 0.0647096335887909
step: 200, loss: 0.017557451501488686
step: 210, loss: 0.0480555035173893
step: 220, loss: 0.001563975471071899
step: 230, loss: 0.006638611666858196
/usr/local/lib/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=0.9898762654668166, f1=0.9875706214689265,
best_f1=0.9875706214689265
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06427494436502457
step: 10, loss: 0.005197259597480297
step: 20, loss: 0.009465742856264114
step: 30, loss: 0.02783312276005745
step: 40, loss: 0.06546641141176224
step: 50, loss: 0.003718102350831032
step: 60, loss: 0.00048730484559200704
step: 70, loss: 0.0008467485313303769
step: 80, loss: 0.005422515328973532
step: 90, loss: 0.002931023482233286
step: 100, loss: 0.0007522017695009708
step: 110, loss: 0.0016115809557959437
step: 120, loss: 0.0014189300127327442
step: 130, loss: 0.0007523534004576504
step: 140, loss: 0.003238721750676632
step: 150, loss: 0.0015333984047174454
step: 160, loss: 0.0008198933792300522
step: 170, loss: 0.033828411251306534
step: 180, loss: 0.0018895373214036226
step: 190, loss: 0.0022612097673118114
step: 200, loss: 0.0014342545764520764
step: 210, loss: 0.02864990569651127
step: 220, loss: 0.0010268958285450935
step: 230, loss: 0.007653180975466967
/usr/local/lib/python3.7/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=0.9888641425389755, f1=0.9855072463768116,
best_f1=0.9875706214689265
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012451934162527323
step: 10, loss: 0.0007021519122645259
step: 20, loss: 0.00035859516356140375
step: 30, loss: 0.0002971513895317912
step: 40, loss: 0.000894000637345016
step: 50, loss: 0.002154015237465501
step: 60, loss: 0.0006359869730658829
step: 70, loss: 0.0027160695753991604
step: 80, loss: 0.050659120082855225
step: 90, loss: 0.008015389554202557
step: 100, loss: 0.003171031130477786

```

```

step: 110, loss: 0.0030452159699052572
step: 120, loss: 0.012186962179839611
step: 130, loss: 0.002226355718448758
step: 140, loss: 0.0009969962993636727
step: 150, loss: 0.0015709289582446218
step: 160, loss: 0.0006112364935688674
step: 170, loss: 0.010004518553614616
step: 180, loss: 0.039059873670339584
step: 190, loss: 0.010404863394796848
step: 200, loss: 0.0004564115952234715
step: 210, loss: 0.0002250144025310874
step: 220, loss: 0.00030610873363912106
step: 230, loss: 0.00021059571008663625
/usr/local/lib/python3.7/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=0.9876819708846584, f1=0.9898762654668166,
best_f1=0.9875706214689265
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028665357967838645
step: 10, loss: 0.00024849639157764614
step: 20, loss: 0.0002481130650267005
step: 30, loss: 0.009673560969531536
step: 40, loss: 0.0003153349971398711
step: 50, loss: 0.0016772800590842962
step: 60, loss: 0.010152734816074371
step: 70, loss: 0.00018417881801724434
step: 80, loss: 0.00023257892462424934
step: 90, loss: 0.00787317380309105
step: 100, loss: 0.0032935384660959244
step: 110, loss: 0.0027924624737352133
step: 120, loss: 0.050330158323049545
step: 130, loss: 0.10646297037601471
step: 140, loss: 0.00026992199127562344
step: 150, loss: 0.0012929342919960618
step: 160, loss: 0.005560979247093201
step: 170, loss: 0.014906731434166431
step: 180, loss: 0.0005698308814316988
step: 190, loss: 0.00032028212444856763
step: 200, loss: 0.001851048320531845
step: 210, loss: 0.001785412197932601
step: 220, loss: 0.00039548747008666396
step: 230, loss: 0.07642212510108948
/usr/local/lib/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=0.9921612541993281, f1=0.9821826280623607,
best_f1=0.9821826280623607
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.029094239696860313
step: 10, loss: 0.0007009789696894586
step: 20, loss: 0.00033809884916990995
step: 30, loss: 0.002606952330097556
step: 40, loss: 0.000508441065903753
step: 50, loss: 0.021512042731046677
step: 60, loss: 0.014640046283602715
step: 70, loss: 0.00018207170069217682
step: 80, loss: 0.0011881105601787567
step: 90, loss: 0.001474892720580101
step: 100, loss: 0.0003395381791051477
step: 110, loss: 0.027657998725771904
step: 120, loss: 0.0002961300779134035
step: 130, loss: 0.0005286273662932217
step: 140, loss: 0.00018780403479468077
step: 150, loss: 0.0001350740494672209
step: 160, loss: 0.006622211541980505
step: 170, loss: 0.009222373366355896
step: 180, loss: 0.00051008106674999
step: 190, loss: 0.0026909199077636003
step: 200, loss: 0.0452297106385231
step: 210, loss: 0.000355962518369779
step: 220, loss: 0.00016731058713048697
step: 230, loss: 0.00137423281557858
/usr/local/lib/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=0.9943883277216611, f1=0.9843749999999999,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022920498624444008
step: 10, loss: 7.978104258654639e-05
step: 20, loss: 8.781161159276962e-05

```

```

step: 30, loss: 0.0001418643514625728
step: 40, loss: 0.00013199145905673504
step: 50, loss: 5.854237315361388e-05
step: 60, loss: 0.0037834502290934324
step: 70, loss: 9.092495020013303e-05
step: 80, loss: 5.991931175231002e-05
step: 90, loss: 7.851218833820894e-05
step: 100, loss: 0.024087313562631607
step: 110, loss: 0.003983627539128065
step: 120, loss: 0.0008285680669359863
step: 130, loss: 0.0004181880212854594
step: 140, loss: 7.803997868904844e-05
step: 150, loss: 0.003323981538414955
step: 160, loss: 0.00021341591491363943
step: 170, loss: 5.884406709810719e-05
step: 180, loss: 0.0020594222005456686
step: 190, loss: 0.000803881383035332
step: 200, loss: 0.11332457512617111
step: 210, loss: 0.0039002648554742336
step: 220, loss: 0.058938492089509964
step: 230, loss: 0.033779632300138474
/usr/local/lib/python3.7/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=0.992108229988726, f1=0.9841986455981941,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09990504384040833
step: 10, loss: 0.017104659229516983
step: 20, loss: 0.001117598614655435
step: 30, loss: 0.0014317060122266412
step: 40, loss: 0.0062746647745370865
step: 50, loss: 0.0018965486669912934
step: 60, loss: 0.005228131543844938
step: 70, loss: 0.0007980926311574876
step: 80, loss: 0.012234168127179146
step: 90, loss: 0.0003886945778504014
step: 100, loss: 0.001044776407070458
step: 110, loss: 0.000839872460346669
step: 120, loss: 0.00016548226994927973
step: 130, loss: 0.005450165364891291
step: 140, loss: 0.0006307667354121804
step: 150, loss: 0.00026759959291666746
step: 160, loss: 0.000361240585334599
step: 170, loss: 0.0016016213921830058
step: 180, loss: 0.0005219315644353628

```

```

step: 190, loss: 0.0029844585806131363
step: 200, loss: 0.0005121569847688079
step: 210, loss: 9.910633525578305e-05
step: 220, loss: 0.00013660179683938622
step: 230, loss: 0.02416612207889557
/usr/local/lib/python3.7/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=0.9921259842519685, f1=0.9798206278026906,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004579216765705496
step: 10, loss: 0.0005198770668357611
step: 20, loss: 0.0008097860845737159
step: 30, loss: 0.00025102708605118096
step: 40, loss: 8.756963507039472e-05
step: 50, loss: 7.902851211838424e-05
step: 60, loss: 0.00011212719982722774
step: 70, loss: 0.0010504450183361769
step: 80, loss: 0.028926359489560127
step: 90, loss: 0.0015750299207866192
step: 100, loss: 0.00011135033855680376
step: 110, loss: 0.00023794974549673498
step: 120, loss: 0.0846451073884964
step: 130, loss: 6.266322452574968e-05
step: 140, loss: 0.0015793938655406237
step: 150, loss: 4.363985135569237e-05
step: 160, loss: 0.00025585159892216325
step: 170, loss: 0.0012485379120334983
step: 180, loss: 0.00022145253024064004
step: 190, loss: 0.00022453595011029392
step: 200, loss: 0.00014777848264202476
step: 210, loss: 0.0003158889012411237
step: 220, loss: 0.027330616489052773
step: 230, loss: 0.014582592993974686
/usr/local/lib/python3.7/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=0.992108229988726, f1=0.9832026875699889,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00039172545075416565
step: 10, loss: 4.4972308387514204e-05
step: 20, loss: 5.292905552778393e-05

```



```

step: 30, loss: 0.0015978582669049501
step: 40, loss: 0.000971597561147064
step: 50, loss: 0.013209318742156029
step: 60, loss: 0.04246797412633896
step: 70, loss: 0.0008276958833448589
step: 80, loss: 0.011460659094154835
step: 90, loss: 0.0001328539801761508
step: 100, loss: 4.0835002437233925e-05
step: 110, loss: 4.38670685980469e-05
step: 120, loss: 0.020749175921082497
step: 130, loss: 7.562149403383955e-05
step: 140, loss: 0.0005865561543032527
step: 150, loss: 3.766896406887099e-05
step: 160, loss: 0.0001666697789914906
step: 170, loss: 5.3879008191870525e-05
step: 180, loss: 0.0069603933952748775
step: 190, loss: 0.00010162546095671132
step: 200, loss: 6.430836947401986e-05
step: 210, loss: 4.3537947931326926e-05
step: 220, loss: 0.020412027835845947
step: 230, loss: 0.0012524549383670092
/usr/local/lib/python3.7/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=0.9909706546275394, f1=0.9820224719101124,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.660092574544251e-05
step: 10, loss: 9.544825297780335e-05
step: 20, loss: 3.541806654538959e-05
step: 30, loss: 0.00017655831470619887
step: 40, loss: 0.00048370545846410096
step: 50, loss: 0.0009590776171535254
step: 60, loss: 0.0004796090361196548
step: 70, loss: 0.0001421319175278768
step: 80, loss: 0.00011853387695737183
step: 90, loss: 0.0001126485803979449
step: 100, loss: 8.027947478694841e-05
step: 110, loss: 7.00520322425291e-05
step: 120, loss: 0.0002388796565355733
step: 130, loss: 2.6460111257620156e-05
step: 140, loss: 0.0003621844807639718
step: 150, loss: 2.9280043236212805e-05
step: 160, loss: 0.0036929238121956587
step: 170, loss: 0.015260016545653343
step: 180, loss: 0.0003133261052425951

```

```

step: 190, loss: 0.00032423451193608344
step: 200, loss: 5.26057010574732e-05
step: 210, loss: 6.56003612675704e-05
step: 220, loss: 5.939036054769531e-05
step: 230, loss: 0.014487252570688725
/usr/local/lib/python3.7/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=0.9899216125419933, f1=0.9788182831661093,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020269275410100818
step: 10, loss: 0.0006263325340114534
step: 20, loss: 5.618002614937723e-05
step: 30, loss: 0.0005217529251240194
step: 40, loss: 3.7623314710799605e-05
step: 50, loss: 0.01534352358430624
step: 60, loss: 0.010920812375843525
step: 70, loss: 6.436675903387368e-05
step: 80, loss: 4.7010500566102564e-05
step: 90, loss: 5.557015902013518e-05
step: 100, loss: 8.248003723565489e-05
step: 110, loss: 0.0005567666958086193
step: 120, loss: 0.0007495744503103197
step: 130, loss: 5.436121136881411e-05
step: 140, loss: 4.5836353820050135e-05
step: 150, loss: 0.0002850904129445553
step: 160, loss: 0.020114785060286522
step: 170, loss: 9.557361045153812e-05
step: 180, loss: 6.900849984958768e-05
step: 190, loss: 0.005952722858637571
step: 200, loss: 0.0674252137541771
step: 210, loss: 7.042471406748518e-05
step: 220, loss: 0.0180482380092144
step: 230, loss: 4.345689740148373e-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=0.992108229988726, f1=0.9831649831649831,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.470614996738732e-05
step: 10, loss: 8.794361929176375e-05
step: 20, loss: 5.0358910812065005e-05

```

```

step: 30, loss: 0.02408849261701107
step: 40, loss: 0.00013739946007262915
step: 50, loss: 3.938609370379709e-05
step: 60, loss: 0.006806537508964539
step: 70, loss: 4.697823169408366e-05
step: 80, loss: 8.308410906465724e-05
step: 90, loss: 2.5223456759704277e-05
step: 100, loss: 3.8737729482818395e-05
step: 110, loss: 4.6040200686547905e-05
step: 120, loss: 4.5936420065118e-05
step: 130, loss: 0.0002193711989093572
step: 140, loss: 0.00015130617248360068
step: 150, loss: 0.01940968818962574
step: 160, loss: 2.3718497686786577e-05
step: 170, loss: 3.0077266274020076e-05
step: 180, loss: 0.000228002667427063
step: 190, loss: 3.569704858819023e-05
step: 200, loss: 4.1049457649933174e-05
step: 210, loss: 3.544120045262389e-05
step: 220, loss: 3.4211283491458744e-05
step: 230, loss: 2.6452617021277547e-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=0.9910112359550561, f1=0.9798657718120806,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.290733729954809e-05
step: 10, loss: 2.1561456378549337e-05
step: 20, loss: 0.016804495826363564
step: 30, loss: 3.49867477780208e-05
step: 40, loss: 2.7137955839862116e-05
step: 50, loss: 3.183537774020806e-05
step: 60, loss: 4.5732129365205765e-05
step: 70, loss: 0.00025895130238495767
step: 80, loss: 3.329192622913979e-05
step: 90, loss: 0.00027395805227570236
step: 100, loss: 0.0023287665098905563
step: 110, loss: 6.343064160319045e-05
step: 120, loss: 4.54137334600091e-05
step: 130, loss: 7.27382066543214e-05
step: 140, loss: 0.0005128239863552153
step: 150, loss: 4.0175538742914796e-05
step: 160, loss: 0.00013858640159014612
step: 170, loss: 2.5521445422782563e-05
step: 180, loss: 3.861400909954682e-05

```

```

step: 190, loss: 0.0004026982933282852
step: 200, loss: 2.484719516360201e-05
step: 210, loss: 0.01672007516026497
step: 220, loss: 3.149600888718851e-05
step: 230, loss: 3.572758942027576e-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=0.990990990990991, f1=0.9820627802690582,
best_f1=0.9843749999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00035183801082894206
step: 10, loss: 5.934791261097416e-05
step: 20, loss: 4.846648153034039e-05
step: 30, loss: 0.0003109580429736525
step: 40, loss: 4.216775414533913e-05
step: 50, loss: 0.00034350581699982285
step: 60, loss: 2.8702139388769865e-05
step: 70, loss: 3.7154375604586676e-05
step: 80, loss: 0.017749574035406113
step: 90, loss: 2.0451545424293727e-05
step: 100, loss: 0.0002476042427588254
step: 110, loss: 3.333626955281943e-05
step: 120, loss: 5.717594831367023e-05
step: 130, loss: 0.0007035896414890885
step: 140, loss: 7.026729872450233e-05
step: 150, loss: 9.425496682524681e-05
step: 160, loss: 0.015488054603338242
step: 170, loss: 2.191547173424624e-05
step: 180, loss: 4.7723075113026425e-05
step: 190, loss: 0.07831845432519913
step: 200, loss: 4.0396822441834956e-05
step: 210, loss: 0.0246072169393301
step: 220, loss: 0.00447838706895709
step: 230, loss: 0.00010368159564677626
/usr/local/lib/python3.7/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=0.990990990990991, f1=0.980963045912654,
best_f1=0.9843749999999999
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, 224.32it/s]
load_f1 = 0.9921787709497207
real_f1 = 0.9921612541993281
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, 245.26it/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: 0.8007302284240723
step: 10, loss: 0.4165038764476776
step: 20, loss: 0.46239280700683594
step: 30, loss: 0.3818768560886383
step: 40, loss: 0.21674230694770813
step: 50, loss: 0.17219769954681396
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.10237781703472137
step: 70, loss: 0.09239554405212402
step: 80, loss: 0.025913873687386513
step: 90, loss: 0.09788689762353897
step: 100, loss: 0.29994696378707886
step: 110, loss: 0.029028674587607384
step: 120, loss: 0.03486613929271698
step: 130, loss: 0.0480746366083622
step: 140, loss: 0.26183950901031494
step: 150, loss: 0.041589319705963135
step: 160, loss: 0.12317988276481628
step: 170, loss: 0.08979017287492752
step: 180, loss: 0.11185042560100555
step: 190, loss: 0.019349634647369385
step: 200, loss: 0.09356596320867538
step: 210, loss: 0.08050289750099182
step: 220, loss: 0.040982019156217575
step: 230, loss: 0.1383548229932785
step: 240, loss: 0.0885637104511261
step: 250, loss: 0.045674026012420654
step: 260, loss: 0.02801750786602497
step: 270, loss: 0.009167535230517387
step: 280, loss: 0.17890392243862152
step: 290, loss: 0.027816543355584145
step: 300, loss: 0.17120224237442017
step: 310, loss: 0.04160211235284805
step: 320, loss: 0.06101026013493538
step: 330, loss: 0.059312496334314346
step: 340, loss: 0.1618737280368805
step: 350, loss: 0.14514921605587006

```

```

step: 360, loss: 0.08239417523145676
step: 370, loss: 0.13664749264717102
step: 380, loss: 0.14273379743099213
step: 390, loss: 0.012021223083138466
step: 400, loss: 0.023824229836463928
step: 410, loss: 0.01974230259656906
step: 420, loss: 0.004368066322058439
step: 430, loss: 0.01563061773777008
step: 440, loss: 0.039867714047431946
step: 450, loss: 0.01802322268486023
step: 460, loss: 0.17706920206546783
step: 470, loss: 0.19737525284290314
step: 480, loss: 0.3113674223423004
step: 490, loss: 0.035168953239917755
step: 500, loss: 0.007763003930449486
step: 510, loss: 0.03407230228185654
step: 520, loss: 0.01494786236435175
step: 530, loss: 0.048263274133205414
/usr/local/lib/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=0.9425815342214057, f1=0.9390746678882271,
best_f1=0.9390746678882271
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06885681301355362
step: 10, loss: 0.14707817137241364
step: 20, loss: 0.1355351209640503
step: 30, loss: 0.10954529047012329
step: 40, loss: 0.005189503077417612
step: 50, loss: 0.13416001200675964
step: 60, loss: 0.20256765186786652
step: 70, loss: 0.08755575120449066
step: 80, loss: 0.009722276590764523
step: 90, loss: 0.0021835824009031057
step: 100, loss: 0.2557787299156189
step: 110, loss: 0.013498381711542606
step: 120, loss: 0.03389231860637665
step: 130, loss: 0.014993016608059406
step: 140, loss: 0.0423428975045681
step: 150, loss: 0.0754796490073204
step: 160, loss: 0.01247295644134283
step: 170, loss: 0.11914403736591339

```



```

step: 180, loss: 0.010243232361972332
step: 190, loss: 0.03816309943795204
step: 200, loss: 0.008099713362753391
step: 210, loss: 0.011787899769842625
step: 220, loss: 0.093080535531044
step: 230, loss: 0.04335658997297287
step: 240, loss: 0.1497819721698761
step: 250, loss: 0.005951716098934412
step: 260, loss: 0.017057213932275772
step: 270, loss: 0.17208510637283325
step: 280, loss: 0.119139663875103
step: 290, loss: 0.07728493958711624
step: 300, loss: 0.08094830811023712
step: 310, loss: 0.16356104612350464
step: 320, loss: 0.030939597636461258
step: 330, loss: 0.01283065602183342
step: 340, loss: 0.009335065260529518
step: 350, loss: 0.04045157879590988
step: 360, loss: 0.017365995794534683
step: 370, loss: 0.0020357612520456314
step: 380, loss: 0.04555921256542206
step: 390, loss: 0.02478322945535183
step: 400, loss: 0.018876725807785988
step: 410, loss: 0.004052064847201109
step: 420, loss: 0.05867470055818558
step: 430, loss: 0.032647907733917236
step: 440, loss: 0.010191376321017742
step: 450, loss: 0.02201058529317379
step: 460, loss: 0.22833041846752167
step: 470, loss: 0.08027656376361847
step: 480, loss: 0.18719464540481567
step: 490, loss: 0.06251882016658783
step: 500, loss: 0.011322171427309513
step: 510, loss: 0.015293908305466175
step: 520, loss: 0.07985905557870865
step: 530, loss: 0.06965843588113785
/usr/local/lib/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=0.943344081068632, f1=0.9387942936033133,
best_f1=0.9387942936033133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.00932224839925766
step: 10, loss: 0.030222494155168533
step: 20, loss: 0.09766598790884018
step: 30, loss: 0.21328110992908478
step: 40, loss: 0.002580679953098297
step: 50, loss: 0.023288756608963013
step: 60, loss: 0.001367664895951748
step: 70, loss: 0.039396364241838455
step: 80, loss: 0.006999166216701269
step: 90, loss: 0.025992583483457565
step: 100, loss: 0.04628749191761017
step: 110, loss: 0.04058777540922165
step: 120, loss: 0.005518000572919846
step: 130, loss: 0.11440305411815643
step: 140, loss: 0.02331097610294819
step: 150, loss: 0.011882390826940536
step: 160, loss: 0.006764184217900038
step: 170, loss: 0.006184721831232309
step: 180, loss: 0.0315348282456398
step: 190, loss: 0.012983650900423527
step: 200, loss: 0.01785930246114731
step: 210, loss: 0.041435882449150085
step: 220, loss: 0.02026287280023098
step: 230, loss: 0.08389419317245483
step: 240, loss: 0.031572625041007996
step: 250, loss: 0.002057342091575265
step: 260, loss: 0.0013446053490042686
step: 270, loss: 0.0034264959394931793
step: 280, loss: 0.0037205354310572147
step: 290, loss: 0.057983480393886566
step: 300, loss: 0.07122036069631577
step: 310, loss: 0.1253148317337036
step: 320, loss: 0.11357807368040085
step: 330, loss: 0.0011001998791471124
step: 340, loss: 0.004434436559677124
step: 350, loss: 0.027572279796004295
step: 360, loss: 0.0037814555689692497
step: 370, loss: 0.0018890504725277424
step: 380, loss: 0.01830136589705944
step: 390, loss: 0.022498631849884987
step: 400, loss: 0.04354117438197136
step: 410, loss: 0.03404692932963371
step: 420, loss: 0.08299192786216736
step: 430, loss: 0.025658953934907913
step: 440, loss: 0.013237224891781807
step: 450, loss: 0.08580555766820908
step: 460, loss: 0.1667013317346573
step: 470, loss: 0.0032640802673995495

```

step: 480, loss: 0.005052921362221241
step: 490, loss: 0.025003300979733467
step: 500, loss: 0.056527383625507355
step: 510, loss: 0.007095005363225937
step: 520, loss: 0.02944292314350605
step: 530, loss: 0.049729324877262115
/usr/local/lib/python3.7/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=0.9372384937238494, f1=0.9346826126954921,
best_f1=0.9387942936033133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021430933848023415
step: 10, loss: 0.009502085857093334
step: 20, loss: 0.09342794120311737
step: 30, loss: 0.013918736018240452
step: 40, loss: 0.001177699537947774
step: 50, loss: 0.02423650212585926
step: 60, loss: 0.0005400493973866105
step: 70, loss: 0.0053192912600934505
step: 80, loss: 0.04156247153878212
step: 90, loss: 0.006562196183949709
step: 100, loss: 0.00342148938216269
step: 110, loss: 0.004827839322388172
step: 120, loss: 0.007677999325096607
step: 130, loss: 0.009257619269192219
step: 140, loss: 0.05247422680258751
step: 150, loss: 0.002829894656315446
step: 160, loss: 0.05863158032298088
step: 170, loss: 0.0029716177377849817
step: 180, loss: 0.009200489148497581
step: 190, loss: 0.08156788349151611
step: 200, loss: 0.0012902385788038373
step: 210, loss: 0.025052886456251144
step: 220, loss: 0.012400874868035316
step: 230, loss: 0.21372923254966736
step: 240, loss: 0.07421614974737167
step: 250, loss: 0.034964390099048615
step: 260, loss: 0.10459432005882263
step: 270, loss: 0.050142280757427216
step: 280, loss: 0.0028464971110224724
step: 290, loss: 0.04786470904946327
step: 300, loss: 0.013281013816595078
step: 310, loss: 0.011555186472833157
step: 320, loss: 0.09013286232948303
step: 330, loss: 0.0532056987285614

```

```

step: 340, loss: 0.06262455135583878
step: 350, loss: 0.002011245349422097
step: 360, loss: 0.004160967655479908
step: 370, loss: 0.042745381593704224
step: 380, loss: 0.0015794463688507676
step: 390, loss: 0.018446212634444237
step: 400, loss: 0.021882152184844017
step: 410, loss: 0.057964541018009186
step: 420, loss: 0.0035131825134158134
step: 430, loss: 0.04976630210876465
step: 440, loss: 0.10788465291261673
step: 450, loss: 0.030427588149905205
step: 460, loss: 0.001057279878295958
step: 470, loss: 0.011254966259002686
step: 480, loss: 0.0024489164352416992
step: 490, loss: 0.012337852269411087
step: 500, loss: 0.010488033294677734
step: 510, loss: 0.017605556175112724
step: 520, loss: 0.025670574977993965
step: 530, loss: 0.0012101909378543496
/usr/local/lib/python3.7/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=0.9411764705882353, f1=0.9395348837209302,
best_f1=0.9387942936033133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016821973025798798
step: 10, loss: 0.0010515230242162943
step: 20, loss: 0.015073906630277634
step: 30, loss: 0.015238218009471893
step: 40, loss: 0.03551677241921425
step: 50, loss: 0.008743231184780598
step: 60, loss: 0.002935457741841674
step: 70, loss: 0.0033770499285310507
step: 80, loss: 0.002757810056209564
step: 90, loss: 0.061859387904405594
step: 100, loss: 0.0009378023096360266
step: 110, loss: 0.002299227751791477
step: 120, loss: 0.003790616989135742
step: 130, loss: 0.0004547186545096338
step: 140, loss: 0.005415001884102821
step: 150, loss: 0.0015108928782865405
step: 160, loss: 0.0244665015488863
step: 170, loss: 0.004584513604640961
step: 180, loss: 0.0011267948430031538
step: 190, loss: 0.0013827835209667683

```

```

step: 200, loss: 0.022188210859894753
step: 210, loss: 0.0006873019738122821
step: 220, loss: 0.00018833132344298065
step: 230, loss: 0.01023892406374216
step: 240, loss: 0.039973873645067215
step: 250, loss: 0.000323449814459309
step: 260, loss: 0.08289922028779984
step: 270, loss: 0.012991349212825298
step: 280, loss: 0.07202423363924026
step: 290, loss: 0.00027023404254578054
step: 300, loss: 0.027603767812252045
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 310, loss: 0.0018409437034279108
step: 320, loss: 0.0037410922814160585
step: 330, loss: 0.007186903152614832
step: 340, loss: 0.003290788037702441
step: 350, loss: 0.0023987689055502415
step: 360, loss: 0.029623093083500862
step: 370, loss: 0.005462591536343098
step: 380, loss: 0.02446194365620613
step: 390, loss: 0.00487699406221509
step: 400, loss: 0.0027788199950009584
step: 410, loss: 0.0012185059022158384
step: 420, loss: 0.0052517796866595745
step: 430, loss: 0.009277886711061
step: 440, loss: 0.0017752725398167968
step: 450, loss: 0.00041235791286453605
step: 460, loss: 0.0025374621618539095
step: 470, loss: 0.07685194164514542
step: 480, loss: 0.01293766126036644
step: 490, loss: 0.0046189455315470695
step: 500, loss: 0.05728878453373909
step: 510, loss: 0.0005781454383395612
step: 520, loss: 0.003257589880377054
step: 530, loss: 0.050968922674655914
/usr/local/lib/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=0.9481000926784059, f1=0.9418819188191883,
best_f1=0.9418819188191883
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004613800439983606

```

step: 10, loss: 0.004069438669830561
step: 20, loss: 0.000536726729478687
step: 30, loss: 0.004753388464450836
step: 40, loss: 0.0010518233757466078
step: 50, loss: 0.0007549934671260417
step: 60, loss: 0.0010782399913296103
step: 70, loss: 0.013433745130896568
step: 80, loss: 0.0076496656984090805
step: 90, loss: 0.034357063472270966
step: 100, loss: 0.11178774386644363
step: 110, loss: 0.08128520101308823
step: 120, loss: 0.0008774286834523082
step: 130, loss: 0.0009165627416223288
step: 140, loss: 0.0004976947675459087
step: 150, loss: 0.01457679271697998
step: 160, loss: 0.0002524301817175001
step: 170, loss: 0.0012166863307356834
step: 180, loss: 0.006675539072602987
step: 190, loss: 0.06546435505151749
step: 200, loss: 0.00045427787699736655
step: 210, loss: 0.0021629291586577892
step: 220, loss: 0.01205152552574873
step: 230, loss: 0.0015554043930023909
step: 240, loss: 0.0002626950736157596
step: 250, loss: 0.01686267927289009
step: 260, loss: 0.013360139913856983
step: 270, loss: 0.00025508637190796435
step: 280, loss: 0.10758879780769348
step: 290, loss: 0.0004621040425263345
step: 300, loss: 0.0027857113163918257
step: 310, loss: 0.0005618595751002431
step: 320, loss: 0.0009654652094468474
step: 330, loss: 0.0012909136712551117
step: 340, loss: 0.0027152744587510824
step: 350, loss: 0.08822912722826004
step: 360, loss: 0.0023363614454865456
step: 370, loss: 0.00030034472001716495
step: 380, loss: 0.13309495151042938
step: 390, loss: 0.014678278006613255
step: 400, loss: 0.0023047367576509714
step: 410, loss: 7.420557813020423e-05
step: 420, loss: 0.061908870935440063
step: 430, loss: 0.00031996957841329277
step: 440, loss: 0.020584773272275925
step: 450, loss: 0.0029116058722138405
step: 460, loss: 0.0028499714098870754
step: 470, loss: 0.05116749182343483
step: 480, loss: 0.008622638881206512

```

step: 490, loss: 0.00010159804514842108
step: 500, loss: 0.01051710732281208
step: 510, loss: 0.001099368673749268
step: 520, loss: 0.016066579148173332
step: 530, loss: 0.01182474009692669
/usr/local/lib/python3.7/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=0.9436619718309859, f1=0.9384109073812883,
best_f1=0.9418819188191883
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05350065603852272
step: 10, loss: 0.030595989897847176
step: 20, loss: 0.056145768612623215
step: 30, loss: 0.0012305257841944695
step: 40, loss: 0.0009372313506901264
step: 50, loss: 0.0005587973864749074
step: 60, loss: 0.12977781891822815
step: 70, loss: 0.01730646751821041
step: 80, loss: 0.0007837365265004337
step: 90, loss: 0.0001363195333397016
step: 100, loss: 0.0005080020637251437
step: 110, loss: 0.0004938106285408139
step: 120, loss: 0.02545161545276642
step: 130, loss: 0.00214094459079206
step: 140, loss: 0.04108191281557083
step: 150, loss: 0.002658916637301445
step: 160, loss: 0.0011145166354253888
step: 170, loss: 0.002861874410882592
step: 180, loss: 0.004510573111474514
step: 190, loss: 0.00550713250413537
step: 200, loss: 0.0018738473299890757
step: 210, loss: 0.0003819834382738918
step: 220, loss: 0.015076713636517525
step: 230, loss: 0.0005054466309957206
step: 240, loss: 0.0033486296888440847
step: 250, loss: 0.001957604195922613
step: 260, loss: 0.0007564620464108884
step: 270, loss: 0.0001270315988222137
step: 280, loss: 0.039139844477176666
step: 290, loss: 0.0025538401678204536
step: 300, loss: 0.0002594792749732733
step: 310, loss: 0.00042651285184547305
step: 320, loss: 0.10288788378238678
step: 330, loss: 0.004888420924544334
step: 340, loss: 0.02928764931857586

```

```
step: 350, loss: 0.00018766475841403008
step: 360, loss: 0.004050802439451218
step: 370, loss: 0.05613351985812187
step: 380, loss: 0.004144828766584396
step: 390, loss: 0.0009994269348680973
step: 400, loss: 6.223675154615194e-05
step: 410, loss: 0.001733087468892336
step: 420, loss: 6.589724944205955e-05
step: 430, loss: 0.00036832134355790913
step: 440, loss: 0.00022159893705975264
step: 450, loss: 0.01344974059611559
step: 460, loss: 0.0011313301511108875
step: 470, loss: 0.10262894630432129
step: 480, loss: 0.003984215669333935
step: 490, loss: 0.0010046145180240273
step: 500, loss: 3.9549126086058095e-05
step: 510, loss: 0.0007024095393717289
step: 520, loss: 0.0008125671884045005
step: 530, loss: 0.00019681658886838704
/usr/local/lib/python3.7/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=0.9443665264142123, f1=0.9399720800372265,
best_f1=0.9418819188191883
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.023792047053575516
step: 10, loss: 0.0037711227778345346
step: 20, loss: 0.020329369232058525
step: 30, loss: 0.037338633090257645
step: 40, loss: 0.00020294327987357974
step: 50, loss: 0.0027194388676434755
step: 60, loss: 0.00010814852430485189
step: 70, loss: 0.0006717818323522806
step: 80, loss: 0.0001358704612357542
step: 90, loss: 0.0014563370496034622
step: 100, loss: 0.0019087346736341715
step: 110, loss: 0.0004040237981826067
step: 120, loss: 0.00012692378368228674
step: 130, loss: 0.014532504603266716
step: 140, loss: 3.574886068236083e-05
step: 150, loss: 0.0007807750371284783
step: 160, loss: 0.061578571796417236
step: 170, loss: 0.03805183246731758
step: 180, loss: 8.537942630937323e-05
step: 190, loss: 0.00924613792449236
step: 200, loss: 0.035247065126895905
```



```

step: 210, loss: 0.00030338403303176165
step: 220, loss: 0.035510819405317307
step: 230, loss: 0.0008435095078311861
step: 240, loss: 0.00011399031791370362
step: 250, loss: 0.010519213043153286
step: 260, loss: 0.025974588468670845
step: 270, loss: 0.00017896003555506468
step: 280, loss: 0.0018010841449722648
step: 290, loss: 0.004074708558619022
step: 300, loss: 0.0008445599814876914
step: 310, loss: 0.033984217792749405
step: 320, loss: 0.0013704828452318907
step: 330, loss: 0.009684589691460133
step: 340, loss: 0.002538834698498249
step: 350, loss: 0.007983392104506493
step: 360, loss: 0.0002868581213988364
step: 370, loss: 0.001163438893854618
step: 380, loss: 0.0008061241824179888
step: 390, loss: 0.0003551311674527824
step: 400, loss: 0.0510554164648056
step: 410, loss: 0.001436317921616137
step: 420, loss: 0.0001359316665912047
step: 430, loss: 8.356105536222458e-05
step: 440, loss: 0.00038700131699442863
step: 450, loss: 0.0008974742377176881
step: 460, loss: 0.00019358197459951043
step: 470, loss: 9.423468873137608e-05
step: 480, loss: 0.0005024197162128985
step: 490, loss: 6.003093585604802e-05
step: 500, loss: 0.009896434843540192
step: 510, loss: 0.00024241623759735376
step: 520, loss: 0.0005924968863837421
step: 530, loss: 0.0004396882723085582
/usr/local/lib/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=0.9495136637332099, f1=0.9456221198156683,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0035739545710384846
step: 10, loss: 0.06699688732624054
step: 20, loss: 0.001206355867907405

```

step: 30, loss: 0.0008657224243506789
step: 40, loss: 0.05729934200644493
step: 50, loss: 0.00018840079428628087
step: 60, loss: 9.228050475940108e-05
step: 70, loss: 0.11619679629802704
step: 80, loss: 0.00015679058560635895
step: 90, loss: 0.0004720657307188958
step: 100, loss: 7.092326268320903e-05
step: 110, loss: 0.001462767831981182
step: 120, loss: 0.0011363336816430092
step: 130, loss: 0.00023626444453839213
step: 140, loss: 0.0006134639261290431
step: 150, loss: 0.0011637276038527489
step: 160, loss: 0.00017997644317802042
step: 170, loss: 0.0002580503933131695
step: 180, loss: 5.4682288464391604e-05
step: 190, loss: 0.000597677833866328
step: 200, loss: 0.0006225080578587949
step: 210, loss: 2.38451666518813e-05
step: 220, loss: 0.0033027075696736574
step: 230, loss: 0.0007221883861348033
step: 240, loss: 0.018840771168470383
step: 250, loss: 0.01408512145280838
step: 260, loss: 0.00044307936332188547
step: 270, loss: 0.004878668114542961
step: 280, loss: 5.581320510827936e-05
step: 290, loss: 9.073434193851426e-05
step: 300, loss: 0.01202585268765688
step: 310, loss: 2.244820279884152e-05
step: 320, loss: 0.0002526088210288435
step: 330, loss: 0.004678778816014528
step: 340, loss: 0.024042179808020592
step: 350, loss: 5.196747224545106e-05
step: 360, loss: 0.026042122393846512
step: 370, loss: 0.0002470881736371666
step: 380, loss: 7.391280814772472e-05
step: 390, loss: 0.0003125223738607019
step: 400, loss: 0.014071673154830933
step: 410, loss: 5.055690780864097e-05
step: 420, loss: 0.0005076868692412972
step: 430, loss: 2.4965976990642957e-05
step: 440, loss: 0.00010037992615252733
step: 450, loss: 0.010111083276569843
step: 460, loss: 4.043010630994104e-05
step: 470, loss: 4.0267805161420256e-05
step: 480, loss: 4.007777533843182e-05
step: 490, loss: 0.0006874529644846916
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 500, loss: 0.0008636477286927402
step: 510, loss: 0.00021763033873867244
step: 520, loss: 0.007655886933207512
step: 530, loss: 0.0005301143391989172
/usr/local/lib/python3.7/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=0.9446529080675422, f1=0.9383177570093458,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005716041661798954
step: 10, loss: 7.545254629803821e-05
step: 20, loss: 0.0021783020347356796
step: 30, loss: 0.0008578971028327942
step: 40, loss: 0.00042324772221036255
step: 50, loss: 0.0002865690621547401
step: 60, loss: 0.0003351321502123028
step: 70, loss: 0.00035614194348454475
step: 80, loss: 0.009084145538508892
step: 90, loss: 0.0034600228536874056
step: 100, loss: 0.016445649787783623
step: 110, loss: 0.0026764515787363052
step: 120, loss: 0.001911912695504725
step: 130, loss: 4.7072207962628454e-05
step: 140, loss: 0.00012363896530587226
step: 150, loss: 0.00030401142430491745
step: 160, loss: 0.00015938826254568994
step: 170, loss: 6.109565583756194e-05
step: 180, loss: 0.0012020032154396176
step: 190, loss: 0.00010881113121286035
step: 200, loss: 0.001014546025544405
step: 210, loss: 0.00031550522544421256
step: 220, loss: 0.00014380172069650143
step: 230, loss: 0.0007112909806892276
step: 240, loss: 0.00026381525094620883
step: 250, loss: 0.001134194782935083
step: 260, loss: 0.02914619632065296
step: 270, loss: 0.0032178289256989956
step: 280, loss: 0.00017054947966244072
step: 290, loss: 3.0039673220016994e-05
step: 300, loss: 0.0028494985308498144
step: 310, loss: 3.58651923306752e-05
step: 320, loss: 0.000894505123142153
step: 330, loss: 0.019518380984663963
step: 340, loss: 0.0002876698272302747
step: 350, loss: 3.3916869142558426e-05
```

```

step: 360, loss: 0.002006221329793334
step: 370, loss: 0.0005240852478891611
step: 380, loss: 0.0028567954432219267
step: 390, loss: 0.0025584970135241747
step: 400, loss: 0.0002988287596963346
step: 410, loss: 0.00016108278941828758
step: 420, loss: 0.006608512718230486
step: 430, loss: 9.613674046704546e-05
step: 440, loss: 4.53034408565145e-05
step: 450, loss: 0.006230120547115803
step: 460, loss: 0.036741919815540314
step: 470, loss: 0.0012081300374120474
step: 480, loss: 0.0006148432730697095
step: 490, loss: 0.025746500119566917
step: 500, loss: 0.014235208742320538
step: 510, loss: 0.000474144151667133
step: 520, loss: 0.0017852229066193104
step: 530, loss: 0.000313073251163587
/usr/local/lib/python3.7/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=0.9435897435897437, f1=0.9406858202038925,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00041505132685415447
step: 10, loss: 0.016427375376224518
step: 20, loss: 0.00012411161151248962
step: 30, loss: 0.0020179515704512596
step: 40, loss: 0.0021650143899023533
step: 50, loss: 0.0015695132315158844
step: 60, loss: 2.8407928766682744e-05
step: 70, loss: 2.1497997295227833e-05
step: 80, loss: 0.0005742786452174187
step: 90, loss: 0.0010238391114398837
step: 100, loss: 0.0008442218531854451
step: 110, loss: 0.0028072884306311607
step: 120, loss: 0.0013747913762927055
step: 130, loss: 0.00038269974174909294
step: 140, loss: 0.0003054721455555409
step: 150, loss: 0.00011128989717690274
step: 160, loss: 0.006760612595826387
step: 170, loss: 0.005823895335197449
step: 180, loss: 6.255571497604251e-05
step: 190, loss: 0.009675472043454647
step: 200, loss: 0.0026042144745588303
step: 210, loss: 3.750623727682978e-05

```

```

step: 220, loss: 0.000541769084520638
step: 230, loss: 0.0001404610084136948
step: 240, loss: 0.00010889094846788794
step: 250, loss: 0.00013885676162317395
step: 260, loss: 3.562374331522733e-05
step: 270, loss: 0.0028812375385314226
step: 280, loss: 0.0005704149953089654
step: 290, loss: 0.07045751810073853
step: 300, loss: 0.09476323425769806
step: 310, loss: 0.006343982648104429
step: 320, loss: 0.018042033538222313
step: 330, loss: 0.0021728493738919497
step: 340, loss: 0.0006910519441589713
step: 350, loss: 0.012511119246482849
step: 360, loss: 0.0010524495737627149
step: 370, loss: 0.0018346509896218777
step: 380, loss: 0.00010986984125338495
step: 390, loss: 0.007224148139357567
step: 400, loss: 9.45184874581173e-05
step: 410, loss: 0.0002701021439861506
step: 420, loss: 0.00024096386914607137
step: 430, loss: 0.00021704328537452966
step: 440, loss: 0.0007192684570327401
step: 450, loss: 4.7088295104913414e-05
step: 460, loss: 0.006700062658637762
step: 470, loss: 0.0030189398676156998
step: 480, loss: 0.0006805917946621776
step: 490, loss: 0.0005276171141304076
step: 500, loss: 6.30277645541355e-05
step: 510, loss: 0.0005830485024489462
step: 520, loss: 1.8350567188463174e-05
step: 530, loss: 3.4539243642939255e-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=0.9433255269320844, f1=0.9435370975268316,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021700566867366433
step: 10, loss: 0.0002154135872842744
step: 20, loss: 1.544864244351629e-05
step: 30, loss: 0.0001120344313676469
step: 40, loss: 0.0009403107105754316
step: 50, loss: 0.004145915620028973
step: 60, loss: 7.807919610058889e-05
step: 70, loss: 0.006312272511422634

```

step: 80, loss: 0.0011506553273648024
step: 90, loss: 0.0004535425396170467
step: 100, loss: 0.0001426613744115457
step: 110, loss: 7.928356353659183e-05
step: 120, loss: 7.706163887633011e-05
step: 130, loss: 4.477867332752794e-05
step: 140, loss: 3.238196950405836e-05
step: 150, loss: 0.006444490049034357
step: 160, loss: 2.7025342205888592e-05
step: 170, loss: 7.361364259850234e-05
step: 180, loss: 0.00014146798639558256
step: 190, loss: 0.0001255833776667714
step: 200, loss: 2.4079405193333514e-05
step: 210, loss: 0.00020038265211042017
step: 220, loss: 0.00039345715777017176
step: 230, loss: 0.0030076107941567898
step: 240, loss: 7.461660425178707e-05
step: 250, loss: 0.00013534717436414212
step: 260, loss: 0.00028258244856260717
step: 270, loss: 0.0010277990950271487
step: 280, loss: 5.871708708582446e-05
step: 290, loss: 0.003546506632119417
step: 300, loss: 0.061347924172878265
step: 310, loss: 0.0006565917865373194
step: 320, loss: 2.70503805950284e-05
step: 330, loss: 0.0016104427631944418
step: 340, loss: 0.009895037859678268
step: 350, loss: 0.017498595640063286
step: 360, loss: 0.0018440301064401865
step: 370, loss: 0.014912216924130917
step: 380, loss: 0.0010370388627052307
step: 390, loss: 2.540896093705669e-05
step: 400, loss: 0.0006919823936186731
step: 410, loss: 0.0034976559691131115
step: 420, loss: 0.010597149841487408
step: 430, loss: 0.00815506186336279
step: 440, loss: 0.01642571948468685
step: 450, loss: 0.0009814107324928045
step: 460, loss: 0.0003450758522376418
step: 470, loss: 0.0007249287446029484
step: 480, loss: 0.0007432799320667982
step: 490, loss: 0.0030011970084160566
step: 500, loss: 0.007464391179382801
step: 510, loss: 0.001857577939517796
step: 520, loss: 0.024770187214016914
step: 530, loss: 0.0007280901190824807
/usr/local/lib/python3.7/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=0.9432029795158287, f1=0.9431870669745959,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001593615161255002
step: 10, loss: 0.00021226696844678372
step: 20, loss: 0.052404679358005524
step: 30, loss: 0.00380861503072083
step: 40, loss: 0.002497375011444092
step: 50, loss: 0.001819472061470151
step: 60, loss: 0.00019686852465383708
step: 70, loss: 0.00014173725503496826
step: 80, loss: 0.00011840726074296981
step: 90, loss: 0.0013781465822830796
step: 100, loss: 1.838757270888891e-05
step: 110, loss: 0.00019697094103321433
step: 120, loss: 0.00018720145453698933
step: 130, loss: 0.0014210863737389445
step: 140, loss: 0.00033632805570960045
step: 150, loss: 6.359774124575779e-05
step: 160, loss: 5.132682417752221e-05
step: 170, loss: 0.0018080900190398097
step: 180, loss: 5.900672113057226e-05
step: 190, loss: 0.0017073474591597915
step: 200, loss: 0.0019187385914847255
step: 210, loss: 0.001345933647826314
step: 220, loss: 0.00018394678772892803
step: 230, loss: 9.227968257619068e-05
step: 240, loss: 0.00011267117224633694
step: 250, loss: 0.04743145778775215
step: 260, loss: 0.0005996382678858936
step: 270, loss: 0.00025169725995510817
step: 280, loss: 1.8149346942664124e-05
step: 290, loss: 0.00180728395935148
step: 300, loss: 0.0075899939984083176
step: 310, loss: 0.003267034189775586
step: 320, loss: 3.597174509195611e-05
step: 330, loss: 0.002296192804351449
step: 340, loss: 4.5387521822704e-05
step: 350, loss: 0.13506542146205902
step: 360, loss: 0.00011821573571069166
step: 370, loss: 2.7152605980518274e-05
step: 380, loss: 0.00011655340495053679
step: 390, loss: 0.0007629336323589087
step: 400, loss: 4.922780499327928e-05
step: 410, loss: 0.00017164958990179002
```

```

step: 420, loss: 0.00013714209489990026
step: 430, loss: 0.002581486478447914
step: 440, loss: 0.0008035392384044826
step: 450, loss: 0.00799169484525919
step: 460, loss: 1.4446380191657227e-05
step: 470, loss: 0.017351651564240456
step: 480, loss: 0.0018690094584599137
step: 490, loss: 0.00020325319201219827
step: 500, loss: 0.00045824903645552695
step: 510, loss: 5.759624400525354e-05
step: 520, loss: 6.06862195127178e-05
step: 530, loss: 0.0017904859269037843
/usr/local/lib/python3.7/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=0.9441903019213174, f1=0.941016333938294,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.348654617089778e-05
step: 10, loss: 0.0009067384526133537
step: 20, loss: 3.853298403555527e-05
step: 30, loss: 0.00011703233030857518
step: 40, loss: 0.006612666416913271
step: 50, loss: 0.0010248575126752257
step: 60, loss: 6.655726610915735e-05
step: 70, loss: 5.427151336334646e-05
step: 80, loss: 0.006815296597778797
step: 90, loss: 1.953491664608009e-05
step: 100, loss: 0.011211002245545387
step: 110, loss: 2.0216362827341072e-05
step: 120, loss: 1.6666643205098808e-05
step: 130, loss: 0.0001677947584539652
step: 140, loss: 0.0543302483856678
step: 150, loss: 0.07302884012460709
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.0013586583081632853
step: 170, loss: 7.983949035406113e-05
step: 180, loss: 0.004242080729454756
step: 190, loss: 0.0003105713112745434
step: 200, loss: 0.027303101494908333
step: 210, loss: 4.966469350620173e-05
step: 220, loss: 0.0004912892472930253
step: 230, loss: 0.0016537043265998363
step: 240, loss: 0.00017209340876433998
step: 250, loss: 3.8084683183114976e-05
step: 260, loss: 1.6730033166822977e-05

```



```

step: 270, loss: 0.007398570422083139
step: 280, loss: 2.9147075110813603e-05
step: 290, loss: 0.00022440188331529498
step: 300, loss: 3.966363146901131e-05
step: 310, loss: 2.349463466089219e-05
step: 320, loss: 0.00027722486993297935
step: 330, loss: 0.0006012549274601042
step: 340, loss: 0.0001039561175275594
step: 350, loss: 0.0015504582552239299
step: 360, loss: 0.002131005050614476
step: 370, loss: 0.00010175436909776181
step: 380, loss: 0.0005977620603516698
step: 390, loss: 0.0025731250643730164
step: 400, loss: 1.55528869072441e-05
step: 410, loss: 6.164373189676553e-05
step: 420, loss: 0.0005651202518492937
step: 430, loss: 0.00010360346641391516
step: 440, loss: 1.3466820746543817e-05
step: 450, loss: 3.3779957448132336e-05
step: 460, loss: 0.00030179033637978137
step: 470, loss: 1.9598381186369807e-05
step: 480, loss: 1.5083540347404778e-05
step: 490, loss: 4.550533776637167e-05
step: 500, loss: 4.444379737833515e-05
step: 510, loss: 0.023858703672885895
step: 520, loss: 1.2691958545474336e-05
step: 530, loss: 4.293125675758347e-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=0.9440298507462688, f1=0.9398704902867714,
best_f1=0.9456221198156683
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.787014429690316e-05
step: 10, loss: 2.4730841687414795e-05
step: 20, loss: 0.0074889762327075005
step: 30, loss: 6.745827704435214e-05
step: 40, loss: 0.00021546310745179653
step: 50, loss: 0.0005531752249225974
step: 60, loss: 1.9594421246438287e-05
step: 70, loss: 0.00010473758447915316
step: 80, loss: 0.004461162257939577
step: 90, loss: 1.3250655683805235e-05
step: 100, loss: 3.854970418615267e-05
step: 110, loss: 0.017557671293616295
step: 120, loss: 0.0013320654397830367

```

```

step: 130, loss: 4.2685849621193483e-05
step: 140, loss: 1.2833506843890063e-05
step: 150, loss: 0.023021895438432693
step: 160, loss: 2.4481883883709088e-05
step: 170, loss: 0.0006512273685075343
step: 180, loss: 0.0016664407448843122
step: 190, loss: 0.0005289132823236287
step: 200, loss: 0.00019012982374988496
step: 210, loss: 0.00013502364163286984
step: 220, loss: 5.597420749836601e-05
step: 230, loss: 7.32528860680759e-05
step: 240, loss: 0.0015759568195790052
step: 250, loss: 0.00014147275942377746
step: 260, loss: 0.0013126676203683019
step: 270, loss: 0.000641885562799871
step: 280, loss: 0.0010221846168860793
step: 290, loss: 0.001805566600523889
step: 300, loss: 0.01245138794183731
step: 310, loss: 0.0004026913084089756
step: 320, loss: 5.96430036239326e-05
step: 330, loss: 6.161551573313773e-05
step: 340, loss: 0.00017897474754136056
step: 350, loss: 6.0029138694517314e-05
step: 360, loss: 3.7780286220368e-05
step: 370, loss: 3.35281110892538e-05
step: 380, loss: 5.046370279160328e-05
step: 390, loss: 0.0003446874616201967
step: 400, loss: 2.6580206395010464e-05
step: 410, loss: 0.0006536616710945964
step: 420, loss: 0.002926431829109788
step: 430, loss: 5.319122283253819e-05
step: 440, loss: 0.03920689597725868
step: 450, loss: 0.007193641271442175
step: 460, loss: 2.165815567423124e-05
step: 470, loss: 0.00024016003590077162
step: 480, loss: 0.0001986887218663469
step: 490, loss: 0.001140502979978919
step: 500, loss: 0.0010537513298913836
step: 510, loss: 7.383638876490295e-05
step: 520, loss: 3.726167051354423e-05
step: 530, loss: 0.000105775281554088
/usr/local/lib/python3.7/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=0.9454378725355341, f1=0.9421939007737824,
best_f1=0.9456221198156683
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, 251.86it/s]

load_f1 = 0.9448082319925164

real_f1 = 0.9431658055425082

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, 247.31it/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/

```

Downloading: 100% 442/442 [00:00<00:00, 384kB/s]

Downloading: 100% 232k/232k [00:00<00:00, 255kB/s]

Downloading: 100% 268M/268M [00:04<00:00, 55.9MB/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: 0.8452678322792053
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.060601044446229935
step: 20, loss: 0.379040002822876
step: 30, loss: 0.3706852197647095
step: 40, loss: 0.4963432848453522
step: 50, loss: 0.2843616008758545
step: 60, loss: 0.3408399224281311
step: 70, loss: 0.2042675018310547
step: 80, loss: 0.3038962483406067
step: 90, loss: 0.3668023347854614
step: 100, loss: 0.09555307030677795
step: 110, loss: 0.2677818238735199
step: 120, loss: 0.17134861648082733
step: 130, loss: 0.1665540635585785
step: 140, loss: 0.13829739391803741
step: 150, loss: 0.3738548159599304
step: 160, loss: 0.2285473495721817
step: 170, loss: 0.04982997104525566
step: 180, loss: 0.15777452290058136
step: 190, loss: 0.21548528969287872
step: 200, loss: 0.14146220684051514
step: 210, loss: 0.37191587686538696
/usr/local/lib/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=0.6613545816733067, f1=0.6935483870967742,
best_f1=0.6935483870967742
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09778086841106415
step: 10, loss: 0.05423468351364136
step: 20, loss: 0.19535458087921143
step: 30, loss: 0.09275665134191513
step: 40, loss: 0.037131864577531815
step: 50, loss: 0.26233458518981934
step: 60, loss: 0.10203826427459717
step: 70, loss: 0.09541468322277069
step: 80, loss: 0.13374170660972595
step: 90, loss: 0.07774033397436142
step: 100, loss: 0.05162379518151283
step: 110, loss: 0.04937371984124184
step: 120, loss: 0.18245786428451538
step: 130, loss: 0.14144481718540192
step: 140, loss: 0.14609681069850922
step: 150, loss: 0.0367162711918354
step: 160, loss: 0.2269090712070465
step: 170, loss: 0.13805872201919556
step: 180, loss: 0.2357618510723114
step: 190, loss: 0.10603810846805573
step: 200, loss: 0.06790884584188461
step: 210, loss: 0.12754671275615692
/usr/local/lib/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=0.7104722792607804, f1=0.6792452830188679,
best_f1=0.6792452830188679
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10595189779996872
step: 10, loss: 0.14335119724273682
step: 20, loss: 0.22609680891036987
step: 30, loss: 0.0808529257774353
step: 40, loss: 0.21568627655506134
step: 50, loss: 0.11715460568666458
step: 60, loss: 0.37486496567726135
step: 70, loss: 0.09223996847867966
step: 80, loss: 0.0498204343020916

```

```

step: 90, loss: 0.15747013688087463
step: 100, loss: 0.045833297073841095
step: 110, loss: 0.09536614269018173
step: 120, loss: 0.193435937166214
step: 130, loss: 0.14296448230743408
step: 140, loss: 0.26889941096305847
step: 150, loss: 0.1907942295074463
step: 160, loss: 0.13436740636825562
step: 170, loss: 0.13805897533893585
step: 180, loss: 0.02031012251973152
step: 190, loss: 0.14899882674217224
step: 200, loss: 0.08450541645288467
step: 210, loss: 0.11636373400688171
/usr/local/lib/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=0.717557251908397,
best_f1=0.717557251908397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08241277933120728
step: 10, loss: 0.036245912313461304
step: 20, loss: 0.10429535806179047
step: 30, loss: 0.04036296531558037
step: 40, loss: 0.0793519839644432
step: 50, loss: 0.053742699325084686
step: 60, loss: 0.05368709936738014
step: 70, loss: 0.2368471771478653
step: 80, loss: 0.08850155770778656
step: 90, loss: 0.010532258078455925
step: 100, loss: 0.14969371259212494
step: 110, loss: 0.04534967243671417
step: 120, loss: 0.017209231853485107
step: 130, loss: 0.15821513533592224
step: 140, loss: 0.05149468779563904
step: 150, loss: 0.2502804100513458
step: 160, loss: 0.04704921320080757
step: 170, loss: 0.0671110600233078
step: 180, loss: 0.22976528108119965
step: 190, loss: 0.0904858410358429
step: 200, loss: 0.018742477521300316
step: 210, loss: 0.024793555960059166
/usr/local/lib/python3.7/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=0.7290076335877863, f1=0.7366412213740458,
best_f1=0.717557251908397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0636109858751297
step: 10, loss: 0.0961512103676796
step: 20, loss: 0.019295278936624527
step: 30, loss: 0.017979349941015244
step: 40, loss: 0.1969854235649109
step: 50, loss: 0.14441289007663727
step: 60, loss: 0.0362323634326458
step: 70, loss: 0.18225106596946716
step: 80, loss: 0.006996569689363241
step: 90, loss: 0.17866474390029907
step: 100, loss: 0.09051210433244705
step: 110, loss: 0.009441228583455086
step: 120, loss: 0.020905546844005585
step: 130, loss: 0.044245846569538116
step: 140, loss: 0.19346916675567627
step: 150, loss: 0.04128012806177139
step: 160, loss: 0.06553228944540024
step: 170, loss: 0.10794714838266373
step: 180, loss: 0.11802259087562561
step: 190, loss: 0.09466686844825745
step: 200, loss: 0.12221221625804901
step: 210, loss: 0.029916927218437195
/usr/local/lib/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=0.769825918762089, f1=0.7475538160469667,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13126787543296814
step: 10, loss: 0.014563710428774357
step: 20, loss: 0.0987057313323021
step: 30, loss: 0.09609420597553253
step: 40, loss: 0.00899119395762682
step: 50, loss: 0.024545280262827873
step: 60, loss: 0.23761971294879913

```



```

step: 70, loss: 0.03149852529168129
step: 80, loss: 0.07527798414230347
step: 90, loss: 0.0575556606054306
step: 100, loss: 0.018612384796142578
step: 110, loss: 0.007217928767204285
step: 120, loss: 0.016746532171964645
step: 130, loss: 0.0035712665412575006
step: 140, loss: 0.032049618661403656
step: 150, loss: 0.02694627083837986
step: 160, loss: 0.02440750226378441
step: 170, loss: 0.0349065400660038
step: 180, loss: 0.012185347266495228
step: 190, loss: 0.13265284895896912
step: 200, loss: 0.028049815446138382
step: 210, loss: 0.010837090201675892
/usr/local/lib/python3.7/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=0.7391304347826088, f1=0.7329434697855752,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03296045958995819
step: 10, loss: 0.03651249781250954
step: 20, loss: 0.031148673966526985
step: 30, loss: 0.04992835223674774
step: 40, loss: 0.11263459175825119
step: 50, loss: 0.13686366379261017
step: 60, loss: 0.29828569293022156
step: 70, loss: 0.048334747552871704
step: 80, loss: 0.05699453130364418
step: 90, loss: 0.001268505584448576
step: 100, loss: 0.009948602877557278
step: 110, loss: 0.04813757166266441
step: 120, loss: 0.09317952394485474
step: 130, loss: 0.008859421126544476
step: 140, loss: 0.0016891268314793706
step: 150, loss: 0.08210232108831406
step: 160, loss: 0.04762670397758484
step: 170, loss: 0.11931432783603668
step: 180, loss: 0.00792379304766655
step: 190, loss: 0.006651703268289566
step: 200, loss: 0.010892711579799652
step: 210, loss: 0.16293549537658691
/usr/local/lib/python3.7/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=0.7279411764705882, f1=0.7204301075268816,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006709664594382048
step: 10, loss: 0.030688494443893433
step: 20, loss: 0.040852416306734085
step: 30, loss: 0.019246533513069153
step: 40, loss: 0.09665440768003464
step: 50, loss: 0.04575691372156143
step: 60, loss: 0.5116384625434875
step: 70, loss: 0.010819651186466217
step: 80, loss: 0.06257762759923935
step: 90, loss: 0.2243870496749878
step: 100, loss: 0.0527152344584465
step: 110, loss: 0.00471634604036808
step: 120, loss: 0.07740112394094467
step: 130, loss: 0.100868359208107
step: 140, loss: 0.0036973869428038597
step: 150, loss: 0.021852592006325722
step: 160, loss: 0.054609451442956924
step: 170, loss: 0.08780030161142349
step: 180, loss: 0.03361668437719345
step: 190, loss: 0.061896611005067825
step: 200, loss: 0.07177260518074036
step: 210, loss: 0.15605901181697845
/usr/local/lib/python3.7/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=0.7265917602996256, f1=0.7142857142857143,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011136753484606743
step: 10, loss: 0.0516376756131649
step: 20, loss: 0.15001143515110016
step: 30, loss: 0.19612231850624084
step: 40, loss: 0.032046929001808167
step: 50, loss: 0.005215838551521301
step: 60, loss: 0.06874282658100128
step: 70, loss: 0.010538163594901562
step: 80, loss: 0.005992247257381678
step: 90, loss: 0.06140028312802315
step: 100, loss: 0.024264534935355186
step: 110, loss: 0.05130138620734215
step: 120, loss: 0.013969046995043755

```

```

step: 130, loss: 0.00502521637827158
step: 140, loss: 0.02084401622414589
step: 150, loss: 0.03145890682935715
step: 160, loss: 0.017729056999087334
step: 170, loss: 0.03746320307254791
step: 180, loss: 0.11094928532838821
step: 190, loss: 0.20262062549591064
step: 200, loss: 0.022476885467767715
step: 210, loss: 0.13417939841747284
/usr/local/lib/python3.7/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=0.7159090909090908, f1=0.7101449275362319,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025917325168848038
step: 10, loss: 0.0020741319749504328
step: 20, loss: 0.03195252642035484
step: 30, loss: 0.01526265125721693
step: 40, loss: 0.24335822463035583
step: 50, loss: 0.04299332574009895
step: 60, loss: 0.08069566637277603
step: 70, loss: 0.008770151995122433
step: 80, loss: 0.0020607137121260166
step: 90, loss: 0.05608867108821869
step: 100, loss: 0.00402914360165596
step: 110, loss: 0.06070781499147415
step: 120, loss: 0.05209044739603996
step: 130, loss: 0.018921075388789177
step: 140, loss: 0.0009868351044133306
step: 150, loss: 0.007201050408184528
step: 160, loss: 0.05065033957362175
step: 170, loss: 0.05270790308713913
step: 180, loss: 0.007512278389185667
step: 190, loss: 0.1101648360490799
step: 200, loss: 0.0449184887111187
step: 210, loss: 0.02187681756913662
/usr/local/lib/python3.7/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=0.7314285714285713, f1=0.7302752293577981,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005093835294246674

```

```

step: 10, loss: 0.13672807812690735
step: 20, loss: 0.005264175124466419
step: 30, loss: 0.04886125400662422
step: 40, loss: 0.05529681220650673
step: 50, loss: 0.13105419278144836
step: 60, loss: 0.06785392016172409
step: 70, loss: 0.014552207663655281
step: 80, loss: 0.14632542431354523
step: 90, loss: 0.026066243648529053
step: 100, loss: 0.0017941102851182222
step: 110, loss: 0.05373179912567139
step: 120, loss: 0.009075398556888103
step: 130, loss: 0.03963993862271309
step: 140, loss: 0.11346134543418884
step: 150, loss: 0.054740943014621735
step: 160, loss: 0.0023674240801483393
step: 170, loss: 0.16212549805641174
step: 180, loss: 0.024430587887763977
step: 190, loss: 0.06989322602748871
step: 200, loss: 0.0004784093762282282
step: 210, loss: 0.004912599455565214
/usr/local/lib/python3.7/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=0.725868725868726, f1=0.7304015296367113,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001603268668986857
step: 10, loss: 0.0025500610936433077
step: 20, loss: 0.0055963811464607716
step: 30, loss: 0.002290809527039528
step: 40, loss: 0.03876076266169548
step: 50, loss: 0.04901454597711563
step: 60, loss: 0.052198972553014755
step: 70, loss: 0.0025248085148632526
step: 80, loss: 0.05907021090388298
step: 90, loss: 0.004203708842396736
step: 100, loss: 0.003248790744692087
step: 110, loss: 0.005559753626585007
step: 120, loss: 0.00862948689609766
step: 130, loss: 0.026942921802401543
step: 140, loss: 0.013887445442378521
step: 150, loss: 0.026145562529563904
step: 160, loss: 0.006123221013695002
step: 170, loss: 0.009601273573935032
step: 180, loss: 0.07135690748691559

```

```

step: 190, loss: 0.16952913999557495
step: 200, loss: 0.08563487976789474
step: 210, loss: 0.009379984810948372
/usr/local/lib/python3.7/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=0.7276190476190476, f1=0.7192660550458716,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017760521732270718
step: 10, loss: 0.004587565548717976
step: 20, loss: 0.0021217274479568005
step: 30, loss: 0.026529720053076744
step: 40, loss: 0.0010130874579772353
step: 50, loss: 0.00207356084138155
step: 60, loss: 0.018930034711956978
step: 70, loss: 0.04971233755350113
step: 80, loss: 0.08865610510110855
step: 90, loss: 0.0022789400536566973
step: 100, loss: 0.08429136127233505
step: 110, loss: 0.045475561171770096
step: 120, loss: 0.006386136170476675
step: 130, loss: 0.004711770452558994
step: 140, loss: 0.12334734946489334
step: 150, loss: 0.0008879987872205675
step: 160, loss: 0.004975362680852413
step: 170, loss: 0.08274537324905396
step: 180, loss: 0.03339248150587082
step: 190, loss: 0.003980915527790785
step: 200, loss: 0.015943411737680435
step: 210, loss: 0.006089447066187859
/usr/local/lib/python3.7/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=0.745697896749522, f1=0.724264705882353,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009505814872682095
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.001107346499338746
step: 20, loss: 0.001671419246122241
step: 30, loss: 0.1188599169254303
step: 40, loss: 0.012688160873949528
step: 50, loss: 0.022599367424845695

```

```

step: 60, loss: 0.016063697636127472
step: 70, loss: 0.04577033594250679
step: 80, loss: 0.06531064212322235
step: 90, loss: 0.035169560462236404
step: 100, loss: 0.0013910389970988035
step: 110, loss: 0.006753414869308472
step: 120, loss: 0.003952363971620798
step: 130, loss: 0.027439238503575325
step: 140, loss: 0.00042114907409995794
step: 150, loss: 0.06675199419260025
step: 160, loss: 0.05806427821516991
step: 170, loss: 0.00361810065805912
step: 180, loss: 0.012547559104859829
step: 190, loss: 0.0036519598215818405
step: 200, loss: 0.010030396282672882
step: 210, loss: 0.10557949542999268
/usr/local/lib/python3.7/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=0.7300380228136882, f1=0.7205882352941176,
best_f1=0.7475538160469667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005098260473459959
step: 10, loss: 0.005709523800760508
step: 20, loss: 0.006113398354500532
step: 30, loss: 0.023261480033397675
step: 40, loss: 0.0012352126650512218
step: 50, loss: 0.004459921736270189
step: 60, loss: 0.018832668662071228
step: 70, loss: 0.02068307437002659
step: 80, loss: 0.009045806713402271
step: 90, loss: 0.01782926544547081
step: 100, loss: 0.0021025328896939754
step: 110, loss: 0.0006859046989120543
step: 120, loss: 0.02158975787460804
step: 130, loss: 0.06154900789260864
step: 140, loss: 0.018330680206418037
step: 150, loss: 0.06520278006792068
step: 160, loss: 0.005992633756250143
step: 170, loss: 0.02203543297946453
step: 180, loss: 0.037356164306402206
step: 190, loss: 0.04530252888798714
step: 200, loss: 0.03267286345362663
step: 210, loss: 0.04085206240415573
/usr/local/lib/python3.7/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=0.7265774378585086, f1=0.724907063197026,
best_f1=0.7475538160469667
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, 347.20it/s]
load_f1 = 0.752851711026616
real_f1 = 0.75
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, 263.64it/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: 0.8582499027252197
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1690657138824463
step: 20, loss: 0.14806652069091797
step: 30, loss: 0.5075734853744507
step: 40, loss: 0.2559688985347748
step: 50, loss: 0.30812686681747437
step: 60, loss: 0.36334386467933655
step: 70, loss: 0.17084957659244537
step: 80, loss: 0.5039627552032471
step: 90, loss: 0.23201408982276917
step: 100, loss: 0.21500034630298615
step: 110, loss: 0.23195022344589233
step: 120, loss: 0.40371471643447876
step: 130, loss: 0.3569832146167755
step: 140, loss: 0.3135315775871277
step: 150, loss: 0.1910797506570816
step: 160, loss: 0.20803430676460266
step: 170, loss: 0.24768656492233276
step: 180, loss: 0.27263352274894714
step: 190, loss: 0.11575241386890411
/usr/local/lib/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=0.7204610951008645, f1=0.7256637168141593,
best_f1=0.7256637168141593
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18777896463871002
step: 10, loss: 0.013293461874127388
step: 20, loss: 0.15615539252758026
step: 30, loss: 0.13103538751602173
step: 40, loss: 0.29910361766815186
step: 50, loss: 0.3271472156047821
step: 60, loss: 0.098777174949646
step: 70, loss: 0.1565808206796646
step: 80, loss: 0.11130127310752869
step: 90, loss: 0.20216557383537292
step: 100, loss: 0.41017186641693115
step: 110, loss: 0.1357680857181549
step: 120, loss: 0.1358550488948822
step: 130, loss: 0.192127525806427
step: 140, loss: 0.06672929972410202
step: 150, loss: 0.03891228884458542
step: 160, loss: 0.013412765227258205
step: 170, loss: 0.21010348200798035
step: 180, loss: 0.10896515846252441
step: 190, loss: 0.11795823276042938
/usr/local/lib/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=0.7966573816155988, f1=0.7582417582417582,
best_f1=0.7582417582417582
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08951276540756226
step: 10, loss: 0.27326998114585876
step: 20, loss: 0.06867006421089172
step: 30, loss: 0.02143244817852974
step: 40, loss: 0.05365868657827377
step: 50, loss: 0.21228429675102234
step: 60, loss: 0.06330883502960205
step: 70, loss: 0.04460064694285393
step: 80, loss: 0.2520733177661896
step: 90, loss: 0.04908100143074989
step: 100, loss: 0.04277003929018974

```

```

step: 110, loss: 0.25876563787460327
step: 120, loss: 0.015093717724084854
step: 130, loss: 0.008660860359668732
step: 140, loss: 0.027705254033207893
step: 150, loss: 0.08931756764650345
step: 160, loss: 0.2031448781490326
step: 170, loss: 0.0943957269191742
step: 180, loss: 0.03559652343392372
step: 190, loss: 0.024914950132369995
/usr/local/lib/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=0.815426997245179, f1=0.7838616714697405,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.058256201446056366
step: 10, loss: 0.02059132046997547
step: 20, loss: 0.035310257226228714
step: 30, loss: 0.006415626499801874
step: 40, loss: 0.0014065660070627928
step: 50, loss: 0.04853377863764763
step: 60, loss: 0.05583776906132698
step: 70, loss: 0.014176787808537483
step: 80, loss: 0.053770165890455246
step: 90, loss: 0.11396633088588715
step: 100, loss: 0.03755619376897812
step: 110, loss: 0.01530836708843708
step: 120, loss: 0.008673010393977165
step: 130, loss: 0.17345955967903137
step: 140, loss: 0.1266225427389145
step: 150, loss: 0.008756850846111774
step: 160, loss: 0.0323166698217392
step: 170, loss: 0.004744069650769234
step: 180, loss: 0.05019994080066681
step: 190, loss: 0.09754694998264313
/usr/local/lib/python3.7/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=0.795969773299748, f1=0.7893333333333333,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06522456556558609
step: 10, loss: 0.012684698216617107
step: 20, loss: 0.12185774743556976
step: 30, loss: 0.07560110092163086
step: 40, loss: 0.018364068120718002
step: 50, loss: 0.0038013909943401814
step: 60, loss: 0.12389472126960754
step: 70, loss: 0.018703483045101166
step: 80, loss: 0.014666645787656307
step: 90, loss: 0.011739998124539852
step: 100, loss: 0.11546001583337784
step: 110, loss: 0.007491315249353647
step: 120, loss: 0.002310721902176738
step: 130, loss: 0.02926703169941902
step: 140, loss: 0.001472476520575583
step: 150, loss: 0.03819655999541283
step: 160, loss: 0.004411195870488882
step: 170, loss: 0.0803321823477745
step: 180, loss: 0.13220739364624023
step: 190, loss: 0.03957004100084305
/usr/local/lib/python3.7/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=0.8054054054054054, f1=0.7679083094555874,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016548020765185356
step: 10, loss: 0.001812875154428184
step: 20, loss: 0.01042102836072445
step: 30, loss: 0.0052526709623634815
step: 40, loss: 0.12013822793960571
step: 50, loss: 0.004816412925720215
step: 60, loss: 0.0017157973488792777
step: 70, loss: 0.042631007730960846
step: 80, loss: 0.16641558706760406
step: 90, loss: 0.02504434622824192
step: 100, loss: 0.057841312140226364
step: 110, loss: 0.04937715828418732
step: 120, loss: 0.01301969587802887
step: 130, loss: 0.017583265900611877
step: 140, loss: 0.003093350911512971
step: 150, loss: 0.01436652336269617
step: 160, loss: 0.022109273821115494
step: 170, loss: 0.014584841206669807
step: 180, loss: 0.04689452797174454

```

```

step: 190, loss: 0.006276968866586685
/usr/local/lib/python3.7/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=0.8033240997229917, f1=0.7922437673130193,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0025120056234300137
step: 10, loss: 0.007598517928272486
step: 20, loss: 0.007784003391861916
step: 30, loss: 0.012465468607842922
step: 40, loss: 0.005775452125817537
step: 50, loss: 0.19424152374267578
step: 60, loss: 0.0009495348203927279
step: 70, loss: 0.0022884896025061607
step: 80, loss: 0.007587215397506952
step: 90, loss: 0.052678100764751434
step: 100, loss: 0.007837831042706966
step: 110, loss: 0.022492358461022377
step: 120, loss: 0.008725268766283989
step: 130, loss: 0.0009382466087117791
step: 140, loss: 0.004685273393988609
step: 150, loss: 0.00462058512493968
step: 160, loss: 0.03721156343817711
step: 170, loss: 0.13045655190944672
step: 180, loss: 0.02028076909482479
step: 190, loss: 0.0059177023358643055
/usr/local/lib/python3.7/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=0.8, f1=0.7807486631016043, best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002830677665770054
step: 10, loss: 0.06921873241662979
step: 20, loss: 0.12581981718540192
step: 30, loss: 0.005785331595689058
step: 40, loss: 0.03455048054456711
step: 50, loss: 0.0034935090225189924
step: 60, loss: 0.0011286316439509392
step: 70, loss: 0.0029838415794074535
step: 80, loss: 0.040073614567518234
step: 90, loss: 0.028377490118145943
step: 100, loss: 0.08479133993387222
step: 110, loss: 0.0018966636853292584

```

```

step: 120, loss: 0.03665926307439804
step: 130, loss: 0.012483587488532066
step: 140, loss: 0.006824741140007973
step: 150, loss: 0.004519890993833542
step: 160, loss: 0.00135994388256222
step: 170, loss: 0.0003132755809929222
step: 180, loss: 0.006580549292266369
step: 190, loss: 0.022545911371707916
/usr/local/lib/python3.7/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=0.8053333333333333, f1=0.7967032967032966,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018694421742111444
step: 10, loss: 0.006087588611990213
step: 20, loss: 0.03569406643509865
step: 30, loss: 0.0034277369268238544
step: 40, loss: 0.00053837476298213
step: 50, loss: 0.0009757202351465821
step: 60, loss: 0.0009229987626895308
step: 70, loss: 0.0007572355098091066
step: 80, loss: 0.004245959687978029
step: 90, loss: 0.05042507126927376
step: 100, loss: 0.004036662634462118
step: 110, loss: 0.015165532007813454
step: 120, loss: 0.0006796465604566038
step: 130, loss: 0.00455227866768837
step: 140, loss: 0.0032775341533124447
step: 150, loss: 0.0732564777135849
step: 160, loss: 0.011928868480026722
step: 170, loss: 0.038672298192977905
step: 180, loss: 0.007251254748553038
step: 190, loss: 0.017270425334572792
/usr/local/lib/python3.7/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=0.8070175438596492, f1=0.7760416666666666,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005572757800109684
step: 10, loss: 0.010480557568371296
step: 20, loss: 0.0016443332424387336
step: 30, loss: 0.0030948901548981667

```

```

step: 40, loss: 0.0034689507447183132
step: 50, loss: 0.0004307171911932528
step: 60, loss: 0.0010588907171040773
step: 70, loss: 0.06448827683925629
step: 80, loss: 0.006039110943675041
step: 90, loss: 0.06458921730518341
step: 100, loss: 0.00144063716288656
step: 110, loss: 0.06439246237277985
step: 120, loss: 0.003677498549222946
step: 130, loss: 0.012603286653757095
step: 140, loss: 0.018963966518640518
step: 150, loss: 0.009584645740687847
step: 160, loss: 0.00912855751812458
step: 170, loss: 0.0035099047236144543
step: 180, loss: 0.002864999230951071
step: 190, loss: 0.0077929310500621796
/usr/local/lib/python3.7/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=0.8134715025906736, f1=0.7978436657681941,
best_f1=0.7838616714697405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003583858488127589
step: 10, loss: 0.0005046752048656344
step: 20, loss: 0.001553065376356244
step: 30, loss: 0.09990230947732925
step: 40, loss: 0.0022869159001857042
step: 50, loss: 0.0007089885184541345
step: 60, loss: 0.001370379701256752
step: 70, loss: 0.00015944668848533183
step: 80, loss: 0.00047406397061422467
step: 90, loss: 0.0035899358335882425
step: 100, loss: 0.0005998627166263759
step: 110, loss: 0.0009936781134456396
step: 120, loss: 0.006044529844075441
step: 130, loss: 0.0005733388825319707
step: 140, loss: 0.0002847655559889972
step: 150, loss: 0.0005828309222124517
step: 160, loss: 0.002516078995540738
step: 170, loss: 0.005077495239675045
step: 180, loss: 0.025422193109989166
step: 190, loss: 0.0019953204318881035
/usr/local/lib/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=0.8232189973614776, f1=0.8097826086956521,
best_f1=0.8097826086956521
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04294028878211975
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.0016654838109388947
step: 20, loss: 0.001371129066683352
step: 30, loss: 0.13310080766677856
step: 40, loss: 0.0016077657928690314
step: 50, loss: 0.001976854633539915
step: 60, loss: 0.0004098722420167178
step: 70, loss: 0.001201503211632371
step: 80, loss: 0.0004245690943207592
step: 90, loss: 0.00114211265463382
step: 100, loss: 0.0018345443531870842
step: 110, loss: 0.0015203104121610522
step: 120, loss: 0.0017966647865250707
step: 130, loss: 0.0016934871673583984
step: 140, loss: 0.0838611051440239
step: 150, loss: 0.00028375681722536683
step: 160, loss: 0.013325831852853298
step: 170, loss: 0.008521348237991333
step: 180, loss: 0.0024070865474641323
step: 190, loss: 0.145773783326149
/usr/local/lib/python3.7/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=0.8195876288659794, f1=0.7989417989417988,
best_f1=0.8097826086956521
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001527005573734641
step: 10, loss: 0.0022830325178802013
step: 20, loss: 0.001704298541881144
step: 30, loss: 0.0021315745543688536
step: 40, loss: 0.0017001185333356261
step: 50, loss: 0.031870629638433456
step: 60, loss: 0.0017069375608116388
step: 70, loss: 0.01338803768157959
step: 80, loss: 0.0012777523370459676
step: 90, loss: 0.007775879465043545
step: 100, loss: 0.1235618069767952
```



```

step: 110, loss: 0.0005686188815161586
step: 120, loss: 0.0063722021877765656
step: 130, loss: 0.0013615156058222055
step: 140, loss: 0.0050089736469089985
step: 150, loss: 0.009180194698274136
step: 160, loss: 0.0015847304603084922
step: 170, loss: 0.001275328453630209
step: 180, loss: 0.05241819843649864
step: 190, loss: 0.00031364080496132374
/usr/local/lib/python3.7/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=0.80719794344473, f1=0.7819148936170213,
best_f1=0.8097826086956521
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056125909090042114
step: 10, loss: 0.0008966429159045219
step: 20, loss: 0.008927881717681885
step: 30, loss: 0.04604019597172737
step: 40, loss: 0.0028247805312275887
step: 50, loss: 0.0007023541838862002
step: 60, loss: 0.0037884272169321775
step: 70, loss: 0.004540667403489351
step: 80, loss: 0.01241764985024929
step: 90, loss: 0.053348831832408905
step: 100, loss: 0.001169291790574789
step: 110, loss: 0.0014186525950208306
step: 120, loss: 0.001581367920152843
step: 130, loss: 0.0004579401866067201
step: 140, loss: 0.08327040821313858
step: 150, loss: 0.0070525724440813065
step: 160, loss: 0.0008470348548144102
step: 170, loss: 0.0050714579410851
step: 180, loss: 0.007377977948635817
step: 190, loss: 0.00035829091211780906
/usr/local/lib/python3.7/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=0.8113695090439276, f1=0.7924528301886792,
best_f1=0.8097826086956521
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018680496141314507
step: 10, loss: 0.0004248963959980756
step: 20, loss: 0.010380210354924202

```

```

step: 30, loss: 0.04670276492834091
step: 40, loss: 0.005148180760443211
step: 50, loss: 0.0018526779022067785
step: 60, loss: 0.001338619040325284
step: 70, loss: 0.0010062960209324956
step: 80, loss: 0.017500873655080795
step: 90, loss: 0.001925943186506629
step: 100, loss: 0.0008470648899674416
step: 110, loss: 0.009587832726538181
step: 120, loss: 0.001474490505643189
step: 130, loss: 0.004102043341845274
step: 140, loss: 0.09394776821136475
step: 150, loss: 0.0005805642576888204
step: 160, loss: 0.04506497457623482
step: 170, loss: 0.002794399857521057
step: 180, loss: 0.0010129279689863324
step: 190, loss: 0.0009377945098094642
/usr/local/lib/python3.7/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=0.8134715025906736, f1=0.7978436657681941,
best_f1=0.8097826086956521
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, 323.86it/s]
load_f1 = 0.6502057613168725
real_f1 = 0.6305220883534136
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 [00:12, 340.55it/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        : 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: 0.8397343158721924

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

step: 10, loss: 0.21862976253032684

step: 20, loss: 0.1461092084646225

step: 30, loss: 0.27427247166633606

step: 40, loss: 0.3162732720375061

step: 50, loss: 0.3800678253173828

```

step: 60, loss: 0.4542151689529419
step: 70, loss: 0.3131946325302124
step: 80, loss: 0.2522185444831848
step: 90, loss: 0.4177488684654236
step: 100, loss: 0.24025507271289825
step: 110, loss: 0.18387643992900848
step: 120, loss: 0.5039105415344238
step: 130, loss: 0.41392648220062256
step: 140, loss: 0.4729369282722473
step: 150, loss: 0.12980276346206665
step: 160, loss: 0.3230196535587311
step: 170, loss: 0.2040686309337616
/usr/local/lib/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=0.5443298969072166, f1=0.54320987654321,
best_f1=0.54320987654321
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.31103837490081787
step: 10, loss: 0.13492907583713531
step: 20, loss: 0.2628493010997772
step: 30, loss: 0.162218376994133
step: 40, loss: 0.15158751606941223
step: 50, loss: 0.2284679263830185
step: 60, loss: 0.0721006691455841
step: 70, loss: 0.2529481053352356
step: 80, loss: 0.08750615268945694
step: 90, loss: 0.18643593788146973
step: 100, loss: 0.13850846886634827
step: 110, loss: 0.12318752706050873
step: 120, loss: 0.03608941286802292
step: 130, loss: 0.06573238968849182
step: 140, loss: 0.11075931042432785
step: 150, loss: 0.16478827595710754
step: 160, loss: 0.16261804103851318
step: 170, loss: 0.06505376845598221
/usr/local/lib/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=0.7927927927927928, f1=0.7739130434782608,
best_f1=0.7739130434782608
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08005400747060776
step: 10, loss: 0.02376379631459713
step: 20, loss: 0.053235381841659546
step: 30, loss: 0.09881037473678589
step: 40, loss: 0.012681563384830952
step: 50, loss: 0.09541650861501694
step: 60, loss: 0.19271089136600494
step: 70, loss: 0.07478104531764984
step: 80, loss: 0.07881170511245728
step: 90, loss: 0.06046559289097786
step: 100, loss: 0.02503553032875061
step: 110, loss: 0.2148786336183548
step: 120, loss: 0.09903953224420547
step: 130, loss: 0.12034476548433304
step: 140, loss: 0.008259110152721405
step: 150, loss: 0.08366890251636505
step: 160, loss: 0.01407246757298708
step: 170, loss: 0.17909017205238342
/usr/local/lib/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=0.7949367088607596, f1=0.78743961352657,
best_f1=0.78743961352657
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06314577162265778
step: 10, loss: 0.17998866736888885
step: 20, loss: 0.028499435633420944
step: 30, loss: 0.15941931307315826
step: 40, loss: 0.008547676727175713
step: 50, loss: 0.013657113537192345
step: 60, loss: 0.09211993962526321
step: 70, loss: 0.033312562853097916
step: 80, loss: 0.06670594215393066
step: 90, loss: 0.046003881841897964
step: 100, loss: 0.018132293596863747
step: 110, loss: 0.04120500385761261

```

```

step: 120, loss: 0.2317592054605484
step: 130, loss: 0.03058163821697235
step: 140, loss: 0.10449369251728058
step: 150, loss: 0.03136476129293442
step: 160, loss: 0.09328806400299072
step: 170, loss: 0.00615502055734396
/usr/local/lib/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=0.8277634961439588, f1=0.8079800498753117,
best_f1=0.8079800498753117
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07269293814897537
step: 10, loss: 0.12237609177827835
step: 20, loss: 0.016611669212579727
step: 30, loss: 0.16242147982120514
step: 40, loss: 0.054516468197107315
step: 50, loss: 0.11862792074680328
step: 60, loss: 0.002862699097022414
step: 70, loss: 0.018193552270531654
step: 80, loss: 0.008892582729458809
step: 90, loss: 0.010165074840188026
step: 100, loss: 0.026919865980744362
step: 110, loss: 0.023244502022862434
step: 120, loss: 0.11794070154428482
step: 130, loss: 0.014553270302712917
step: 140, loss: 0.041189368814229965
step: 150, loss: 0.012232326902449131
step: 160, loss: 0.019009295850992203
step: 170, loss: 0.09835132956504822
/usr/local/lib/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=0.837092731829574, f1=0.8104265402843601,
best_f1=0.8104265402843601
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.02715890109539032
step: 10, loss: 0.07149249315261841
step: 20, loss: 0.05123163014650345
step: 30, loss: 0.020689884200692177
step: 40, loss: 0.0017825895920395851
step: 50, loss: 0.056567106395959854
step: 60, loss: 0.03519687056541443
step: 70, loss: 0.009495875798165798
step: 80, loss: 0.017486026510596275
step: 90, loss: 0.03154963254928589
step: 100, loss: 0.0611473023891449
step: 110, loss: 0.016835790127515793
step: 120, loss: 0.059067122638225555
step: 130, loss: 0.18276257812976837
step: 140, loss: 0.11640579253435135
step: 150, loss: 0.12996506690979004
step: 160, loss: 0.013266365975141525
step: 170, loss: 0.006091257557272911
/usr/local/lib/python3.7/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=0.8302872062663185, f1=0.8333333333333333,
best_f1=0.8104265402843601
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10056571662425995
step: 10, loss: 0.0033389918971806765
step: 20, loss: 0.029740193858742714
step: 30, loss: 0.008142851293087006
step: 40, loss: 0.021777288988232613
step: 50, loss: 0.0037898337468504906
step: 60, loss: 0.19926677644252777
step: 70, loss: 0.0053030657581985
step: 80, loss: 0.0022503354120999575
step: 90, loss: 0.0098407082259655
step: 100, loss: 0.02066468633711338
step: 110, loss: 0.02394956536591053
step: 120, loss: 0.18613959848880768
step: 130, loss: 0.3133268356323242
step: 140, loss: 0.026020808145403862
step: 150, loss: 0.12974965572357178
step: 160, loss: 0.013720797374844551
step: 170, loss: 0.20469598472118378
/usr/local/lib/python3.7/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=0.835978835978836, f1=0.8258706467661691,

```



```

best_f1=0.8104265402843601
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04057283699512482
step: 10, loss: 0.003749881172552705
step: 20, loss: 0.017133226618170738
step: 30, loss: 0.010359418578445911
step: 40, loss: 0.0014503923011943698
step: 50, loss: 0.003920409362763166
step: 60, loss: 0.003057562978938222
step: 70, loss: 0.024368032813072205
step: 80, loss: 0.013617150485515594
step: 90, loss: 0.016059013083577156
step: 100, loss: 0.0037956428714096546
step: 110, loss: 0.04147050529718399
step: 120, loss: 0.012957721017301083
step: 130, loss: 0.0018024626187980175
step: 140, loss: 0.08844387531280518
step: 150, loss: 0.01540672592818737
step: 160, loss: 0.043144673109054565
step: 170, loss: 0.03323085233569145
/usr/local/lib/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=0.8423772609819121, f1=0.825, best_f1=0.825
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019062621518969536
step: 10, loss: 0.16856077313423157
step: 20, loss: 0.00512394355610013
step: 30, loss: 0.039693258702754974
step: 40, loss: 0.03290661424398422
step: 50, loss: 0.009818712249398232
step: 60, loss: 0.06595960259437561
step: 70, loss: 0.007208103314042091
step: 80, loss: 0.11796824634075165
step: 90, loss: 0.003816802753135562
step: 100, loss: 0.04807066172361374
step: 110, loss: 0.01135107222944498
step: 120, loss: 0.013652865774929523
step: 130, loss: 0.007251512259244919
step: 140, loss: 0.015102984383702278

```

```

step: 150, loss: 0.09630026668310165
step: 160, loss: 0.13316267728805542
step: 170, loss: 0.029455147683620453
/usr/local/lib/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=0.8527131782945736, f1=0.8215158924205379,
best_f1=0.8215158924205379
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002859322587028146
step: 10, loss: 0.03452759608626366
step: 20, loss: 0.0004900008207187057
step: 30, loss: 0.0006986749358475208
step: 40, loss: 0.014803534373641014
step: 50, loss: 0.008442881517112255
step: 60, loss: 0.009187106974422932
step: 70, loss: 0.12981027364730835
step: 80, loss: 0.015854232013225555
step: 90, loss: 0.021357707679271698
step: 100, loss: 0.0002964879386126995
step: 110, loss: 0.019066650420427322
step: 120, loss: 0.02694075182080269
step: 130, loss: 0.03805413469672203
step: 140, loss: 0.01883239671587944
step: 150, loss: 0.0073144338093698025
step: 160, loss: 0.0028796440456062555
step: 170, loss: 0.02307792752981186
/usr/local/lib/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=0.8535353535353535, f1=0.8232445520581113,
best_f1=0.8232445520581113
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002344693202758208
step: 10, loss: 0.030763544142246246
step: 20, loss: 0.04664686694741249

```

```

step: 30, loss: 0.002827390795573592
step: 40, loss: 0.020704636350274086
step: 50, loss: 0.007820784114301205
step: 60, loss: 0.0030775684863328934
step: 70, loss: 0.12422797083854675
step: 80, loss: 0.0012562087504193187
step: 90, loss: 0.00559582794085145
step: 100, loss: 0.0012931609526276588
step: 110, loss: 0.0013879331527277827
step: 120, loss: 0.02712763100862503
step: 130, loss: 0.017024630680680275
step: 140, loss: 0.08477192372083664
step: 150, loss: 0.01517638098448515
step: 160, loss: 0.00441715819761157
step: 170, loss: 0.08457133919000626
/usr/local/lib/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=0.8592233009708737, f1=0.8275862068965516,
best_f1=0.8275862068965516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012646013870835304
step: 10, loss: 0.015086481347680092
step: 20, loss: 0.04727998003363609
step: 30, loss: 0.0006673365132883191
step: 40, loss: 0.0008784335805103183
step: 50, loss: 0.003115813946351409
step: 60, loss: 0.0003542364574968815
step: 70, loss: 0.045962873846292496
step: 80, loss: 0.06746537238359451
step: 90, loss: 0.0008579161367379129
step: 100, loss: 0.0005751103744842112
step: 110, loss: 0.0007121581002138555
step: 120, loss: 0.028233759105205536
step: 130, loss: 0.016146762296557426
step: 140, loss: 0.003909306135028601
step: 150, loss: 0.0025191495660692453
step: 160, loss: 0.007460490800440311
step: 170, loss: 0.02327865920960903
/usr/local/lib/python3.7/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=0.85785536159601, f1=0.8329411764705882,
best_f1=0.8275862068965516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009848092682659626
step: 10, loss: 0.05167747288942337
step: 20, loss: 0.0005084808217361569
step: 30, loss: 0.0048796008341014385
step: 40, loss: 0.2381206899881363
step: 50, loss: 0.0016325847245752811
step: 60, loss: 0.00840072426944971
step: 70, loss: 0.01688375324010849
step: 80, loss: 0.0031794884707778692
step: 90, loss: 0.001350560924038291
step: 100, loss: 0.0010181651450693607
step: 110, loss: 0.05113377794623375
step: 120, loss: 0.0033290409483015537
step: 130, loss: 0.0440320149064064
step: 140, loss: 0.0008199206204153597
step: 150, loss: 0.00028661973192356527
step: 160, loss: 0.004626138601452112
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 170, loss: 0.004963666666299105
/usr/local/lib/python3.7/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=0.8480392156862744, f1=0.8259860788863109,
best_f1=0.8275862068965516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018317296635359526
step: 10, loss: 0.00021816737717017531
step: 20, loss: 0.020903274416923523
step: 30, loss: 0.0002386747219134122
step: 40, loss: 0.007684789597988129
step: 50, loss: 0.024239568039774895
step: 60, loss: 0.0006141945486888289
step: 70, loss: 0.02206401526927948
step: 80, loss: 0.001791269169189036
step: 90, loss: 0.00215320335701108
step: 100, loss: 0.0003028550709132105
step: 110, loss: 0.00730900326743722
step: 120, loss: 0.12222083657979965
step: 130, loss: 0.0011854843469336629
step: 140, loss: 0.025988128036260605
step: 150, loss: 0.0010067480616271496

```

```

step: 160, loss: 0.026877691969275475
step: 170, loss: 0.0008804976823739707
/usr/local/lib/python3.7/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=0.8585365853658536, f1=0.8314087759815242,
best_f1=0.8275862068965516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014756487682461739
step: 10, loss: 0.006891339085996151
step: 20, loss: 0.005036661867052317
step: 30, loss: 0.007417698856443167
step: 40, loss: 0.00249951402656734
step: 50, loss: 0.005590739194303751
step: 60, loss: 0.00032313380506820977
step: 70, loss: 0.007061300799250603
step: 80, loss: 0.0012897023698315024
step: 90, loss: 0.0002343985252082348
step: 100, loss: 0.0009254305041395128
step: 110, loss: 0.00038063921965658665
step: 120, loss: 0.04230586066842079
step: 130, loss: 0.0023757622111588717
step: 140, loss: 0.00947779044508934
step: 150, loss: 0.0632457435131073
step: 160, loss: 0.0002966995525639504
step: 170, loss: 0.004010563250631094
/usr/local/lib/python3.7/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=0.8585365853658536, f1=0.8259860788863109,
best_f1=0.8275862068965516
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 [00:04, 426.13it/s]
load_f1 = 0.5638297872340425
real_f1 = 0.5463743676222598
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 [00:13, 334.37it/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: 0.804879367351532
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4706200361251831
step: 20, loss: 0.6144745945930481
step: 30, loss: 0.4389041066169739
step: 40, loss: 0.15584403276443481
step: 50, loss: 0.3044969439506531
step: 60, loss: 0.17606811225414276
step: 70, loss: 0.15615305304527283
step: 80, loss: 0.21601629257202148
step: 90, loss: 0.08080960065126419
step: 100, loss: 0.2757876217365265
step: 110, loss: 0.10683431476354599
step: 120, loss: 0.07932404428720474
step: 130, loss: 0.028805358335375786
step: 140, loss: 0.004761948250234127
step: 150, loss: 0.15463371574878693
step: 160, loss: 0.13518725335597992
step: 170, loss: 0.006735478062182665
step: 180, loss: 0.0076294089667499065
step: 190, loss: 0.03204246237874031
step: 200, loss: 0.005096801556646824
step: 210, loss: 0.02954552322626114
step: 220, loss: 0.008740170858800411
step: 230, loss: 0.011670469306409359
/usr/local/lib/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=0.9692982456140351, f1=0.9690949227373068,
best_f1=0.9690949227373068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014491410925984383
step: 10, loss: 0.0015517778228968382
step: 20, loss: 0.003126188414171338
step: 30, loss: 0.0022833929397165775
step: 40, loss: 0.0073346300050616264
step: 50, loss: 0.012411362491548061
step: 60, loss: 0.007371311541646719

```



```

step: 70, loss: 0.021719716489315033
step: 80, loss: 0.020566243678331375
step: 90, loss: 0.007232691161334515
step: 100, loss: 0.13562528789043427
step: 110, loss: 0.14078442752361298
step: 120, loss: 0.007223319727927446
step: 130, loss: 0.016349654644727707
step: 140, loss: 0.21073435246944427
step: 150, loss: 0.006716658361256123
step: 160, loss: 0.007235417142510414
step: 170, loss: 0.008311998099088669
step: 180, loss: 0.005394954700022936
step: 190, loss: 0.1359894722700119
step: 200, loss: 0.013885391876101494
step: 210, loss: 0.09189912676811218
step: 220, loss: 0.00491704698652029
step: 230, loss: 0.03918109089136124
/usr/local/lib/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=0.9800443458980044, f1=0.9679558011049725,
best_f1=0.9679558011049725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13679197430610657
step: 10, loss: 0.04192798212170601
step: 20, loss: 0.011653450317680836
step: 30, loss: 0.01808217540383339
step: 40, loss: 0.004561375826597214
step: 50, loss: 0.008336478844285011
step: 60, loss: 0.07966408878564835
step: 70, loss: 0.008587309159338474
step: 80, loss: 0.002355736680328846
step: 90, loss: 0.09293446689844131
step: 100, loss: 0.0021389611065387726
step: 110, loss: 0.004464393947273493
step: 120, loss: 0.01705055683851242
step: 130, loss: 0.0028822373133152723
step: 140, loss: 0.00213965168222785
step: 150, loss: 0.007412688806653023
step: 160, loss: 0.003890948835760355
step: 170, loss: 0.010349864140152931
step: 180, loss: 0.0009274029871448874

```

```

step: 190, loss: 0.02769433706998825
step: 200, loss: 0.0037493363488465548
step: 210, loss: 0.016119636595249176
step: 220, loss: 0.012545231729745865
step: 230, loss: 0.08936748653650284
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.9686800894854586,
best_f1=0.9679558011049725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005698743276298046
step: 10, loss: 0.0031488598324358463
step: 20, loss: 0.0009330944740213454
step: 30, loss: 0.03616729751229286
step: 40, loss: 0.015271907672286034
step: 50, loss: 0.005012770649045706
step: 60, loss: 0.007275963667780161
step: 70, loss: 0.03280086815357208
step: 80, loss: 0.11035232245922089
step: 90, loss: 0.007450674194842577
step: 100, loss: 0.00158622395247221
step: 110, loss: 0.07762310653924942
step: 120, loss: 0.024182533845305443
step: 130, loss: 0.03824576735496521
step: 140, loss: 0.0005213378462940454
step: 150, loss: 0.0019013950368389487
step: 160, loss: 0.01094422210007906
step: 170, loss: 0.0026624954771250486
step: 180, loss: 0.0767773762345314
step: 190, loss: 0.0038488847203552723
step: 200, loss: 0.001917366636916995
step: 210, loss: 0.002582823159173131
step: 220, loss: 0.0031601712107658386
step: 230, loss: 0.00039348937571048737
/usr/local/lib/python3.7/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=0.9776286353467561, f1=0.9765886287625419,
best_f1=0.9679558011049725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005763519438914955
step: 10, loss: 0.001118660205975175
step: 20, loss: 0.0024901542346924543

```

```

step: 30, loss: 0.00147688714787364
step: 40, loss: 0.0002981348952744156
step: 50, loss: 0.0005409705336205661
step: 60, loss: 0.0008174091344699264
step: 70, loss: 0.0018915161490440369
step: 80, loss: 0.0022079378832131624
step: 90, loss: 0.0015357474330812693
step: 100, loss: 0.0010856080334633589
step: 110, loss: 0.005683408118784428
step: 120, loss: 0.04527566581964493
step: 130, loss: 0.0829315185546875
step: 140, loss: 0.012539098039269447
step: 150, loss: 0.0005987719050608575
step: 160, loss: 0.17045044898986816
step: 170, loss: 0.06808619201183319
step: 180, loss: 0.003656249027699232
step: 190, loss: 0.0011989258928224444
step: 200, loss: 0.0014991769567131996
step: 210, loss: 0.0007323483005166054
step: 220, loss: 0.000867652241140604
step: 230, loss: 0.0007491282885894179
/usr/local/lib/python3.7/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=0.9799554565701558, f1=0.9744160177975528,
best_f1=0.9679558011049725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007975093321874738
step: 10, loss: 0.0002540511195547879
step: 20, loss: 0.0005216326680965722
step: 30, loss: 0.0018757589859887958
step: 40, loss: 0.003771125338971615
step: 50, loss: 0.17837800085544586
step: 60, loss: 0.018854135647416115
step: 70, loss: 0.00547035550698638
step: 80, loss: 0.00436755595728755
step: 90, loss: 0.02230057492852211
step: 100, loss: 0.001904840930365026
step: 110, loss: 0.0322074368596077
step: 120, loss: 0.0007511343574151397
step: 130, loss: 0.0018925005570054054
step: 140, loss: 0.0006239543436095119
step: 150, loss: 0.00040238394285552204
step: 160, loss: 0.06392410397529602
step: 170, loss: 0.00014906172873452306
step: 180, loss: 0.0012771347537636757

```

```

step: 190, loss: 0.0002435054921079427
step: 200, loss: 0.06813745200634003
step: 210, loss: 0.007311939261853695
step: 220, loss: 0.0010564608965069056
step: 230, loss: 0.0004615577345248312
/usr/local/lib/python3.7/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=0.9798657718120806, f1=0.9754464285714286,
best_f1=0.9679558011049725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020844101905822754
step: 10, loss: 0.0006141707999631763
step: 20, loss: 0.0066300504840910435
step: 30, loss: 0.0003094340208917856
step: 40, loss: 0.0004526631091721356
step: 50, loss: 8.261146285803989e-05
step: 60, loss: 0.014103717170655727
step: 70, loss: 0.00048637951840646565
step: 80, loss: 0.00024258445773739368
step: 90, loss: 0.000891713541932404
step: 100, loss: 0.0007963576936163008
step: 110, loss: 0.009134537540376186
step: 120, loss: 0.002388505032286048
step: 130, loss: 0.00631230603903532
step: 140, loss: 9.882837912300602e-05
step: 150, loss: 0.013790585100650787
step: 160, loss: 9.039022552315146e-05
step: 170, loss: 0.0005577910342253745
step: 180, loss: 0.0005202361498959363
step: 190, loss: 0.004148357082158327
step: 200, loss: 0.012542295269668102
step: 210, loss: 8.38226405903697e-05
step: 220, loss: 9.328393934993073e-05
step: 230, loss: 0.02702547423541546
/usr/local/lib/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=0.9808773903262092, f1=0.9764837625979844,
best_f1=0.9764837625979844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016322240233421326
step: 10, loss: 0.14120616018772125
step: 20, loss: 0.0001717116974759847
step: 30, loss: 0.00036510571953840554
step: 40, loss: 0.007113747298717499
step: 50, loss: 0.0037233587354421616
step: 60, loss: 0.0006805765442550182
step: 70, loss: 0.00037858591531403363
step: 80, loss: 0.002861435292288661
step: 90, loss: 0.0002498931426089257
step: 100, loss: 0.009833029471337795
step: 110, loss: 0.00023175484966486692
step: 120, loss: 9.713094186736271e-05
step: 130, loss: 0.0023504283744841814
step: 140, loss: 0.0071207936853170395
step: 150, loss: 0.00018190877744928002
step: 160, loss: 0.0008357428596355021
step: 170, loss: 0.007005029357969761
step: 180, loss: 0.0005671855760738254
step: 190, loss: 7.369550439761952e-05
step: 200, loss: 0.01233232207596302
step: 210, loss: 0.0055006868205964565
step: 220, loss: 0.00012762367259711027
step: 230, loss: 0.05086599662899971
/usr/local/lib/python3.7/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=0.978675645342312, f1=0.9720670391061451,
best_f1=0.9764837625979844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006534484564326704
step: 10, loss: 0.0007931820582598448
step: 20, loss: 0.002667628927156329
step: 30, loss: 0.00011616689880611375
step: 40, loss: 0.002643367974087596
step: 50, loss: 6.67460190015845e-05
step: 60, loss: 0.0004840190813411027
step: 70, loss: 0.03241058066487312
step: 80, loss: 0.029607925564050674
step: 90, loss: 0.014017502777278423
step: 100, loss: 0.027681391686201096
step: 110, loss: 0.00013894202129449695
step: 120, loss: 0.0606132373213768
step: 130, loss: 8.974400407169014e-05
step: 140, loss: 0.15099990367889404

```

```

step: 150, loss: 0.00011137937690364197
step: 160, loss: 0.0008537288522347808
step: 170, loss: 0.00019936953322030604
step: 180, loss: 0.0036083078011870384
step: 190, loss: 0.00023834970488678664
step: 200, loss: 0.000312135525746271
step: 210, loss: 9.34399213292636e-05
step: 220, loss: 0.023324621841311455
step: 230, loss: 0.10876000672578812
/usr/local/lib/python3.7/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=0.9787709497206705, f1=0.9711111111111111,
best_f1=0.9764837625979844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012777572264894843
step: 10, loss: 9.096760913962498e-05
step: 20, loss: 0.00014884103438816965
step: 30, loss: 0.0003423921880312264
step: 40, loss: 0.00011387374979676679
step: 50, loss: 0.0007595165516249835
step: 60, loss: 0.047753024846315384
step: 70, loss: 0.0002007630973821506
step: 80, loss: 0.0005670078098773956
step: 90, loss: 9.182385838357732e-05
step: 100, loss: 0.001168811577372253
step: 110, loss: 0.0002573541132733226
step: 120, loss: 0.027135364711284637
step: 130, loss: 0.0035001193173229694
step: 140, loss: 0.025416504591703415
step: 150, loss: 5.1067010645056143e-05
step: 160, loss: 0.00011738827015506104
step: 170, loss: 0.00020154219237156212
step: 180, loss: 0.00022133560560178012
step: 190, loss: 0.0005062497803010046
step: 200, loss: 5.016471550334245e-05
step: 210, loss: 5.4204767366172746e-05
step: 220, loss: 0.00018948836077470332
step: 230, loss: 7.878919859649613e-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=0.9787709497206705, f1=0.9709821428571428,
best_f1=0.9764837625979844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.798973532160744e-05
step: 10, loss: 0.0007284158491529524
step: 20, loss: 3.0155215426930226e-05
step: 30, loss: 0.00020132618374191225
step: 40, loss: 0.0022804178297519684
step: 50, loss: 0.002204306423664093
step: 60, loss: 0.0002823960385285318
step: 70, loss: 0.0013758176937699318
step: 80, loss: 0.00012938599684275687
step: 90, loss: 0.0001278591953450814
step: 100, loss: 0.0014112413628026843
step: 110, loss: 0.008123474195599556
step: 120, loss: 0.0022940882481634617
step: 130, loss: 0.00033477405668236315
step: 140, loss: 6.480397132690996e-05
step: 150, loss: 8.691673428984359e-05
step: 160, loss: 6.965817738091573e-05
step: 170, loss: 0.0177416130900383
step: 180, loss: 0.0012327877338975668
step: 190, loss: 8.395958866458386e-05
step: 200, loss: 7.38229718990624e-05
step: 210, loss: 4.423885911819525e-05
step: 220, loss: 4.370631359051913e-05
step: 230, loss: 0.00025863616610877216
/usr/local/lib/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=0.9821029082774049, f1=0.9700332963374029,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017128491774201393
step: 10, loss: 0.000838848645798862
step: 20, loss: 0.0001539786608191207
step: 30, loss: 0.0004990795277990401
step: 40, loss: 5.099266127217561e-05
step: 50, loss: 0.00024333139299415052
step: 60, loss: 0.003733126213774085
step: 70, loss: 6.581730849575251e-05
step: 80, loss: 3.20625003951136e-05
step: 90, loss: 6.459298310801387e-05
step: 100, loss: 5.396900087362155e-05

```

```

step: 110, loss: 0.0015393306966871023
step: 120, loss: 9.820169361773878e-05
step: 130, loss: 0.0040400344878435135
step: 140, loss: 4.666128370445222e-05
step: 150, loss: 4.481762880459428e-05
step: 160, loss: 0.0416695736348629
step: 170, loss: 6.464920443249866e-05
step: 180, loss: 0.0001375061401631683
step: 190, loss: 0.00010092384036397561
step: 200, loss: 0.011659647338092327
step: 210, loss: 5.526887980522588e-05
step: 220, loss: 0.027896931394934654
step: 230, loss: 8.649544179206714e-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=0.9765363128491621, f1=0.9732142857142857,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.068126771831885e-05
step: 10, loss: 0.004132060334086418
step: 20, loss: 0.00018146095681004226
step: 30, loss: 0.018922390416264534
step: 40, loss: 0.00010085867688758299
step: 50, loss: 8.187410276150331e-05
step: 60, loss: 7.619339157827199e-05
step: 70, loss: 6.0308837419142947e-05
step: 80, loss: 6.359653343679383e-05
step: 90, loss: 6.669369759038091e-05
step: 100, loss: 5.690102989319712e-05
step: 110, loss: 8.56123588164337e-05
step: 120, loss: 0.00012753228656947613
step: 130, loss: 0.0003509125963319093
step: 140, loss: 0.10177398473024368
step: 150, loss: 0.01969865895807743
step: 160, loss: 0.0003530856338329613
step: 170, loss: 6.819504778832197e-05
step: 180, loss: 0.0005590706714428961
step: 190, loss: 7.540914521086961e-05
step: 200, loss: 8.427516877418384e-05
step: 210, loss: 9.326613508164883e-05
step: 220, loss: 0.00042063367436639965
step: 230, loss: 4.5004104322288185e-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=0.9807909604519773, f1=0.9752808988764046,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013955312897451222
step: 10, loss: 5.5939894082257524e-05
step: 20, loss: 0.01708144135773182
step: 30, loss: 0.00978842843323946
step: 40, loss: 3.891560118063353e-05
step: 50, loss: 0.0012823289725929499
step: 60, loss: 0.0001988446165341884
step: 70, loss: 8.926396549213678e-05
step: 80, loss: 0.00020647210476454347
step: 90, loss: 0.0002407289866823703
step: 100, loss: 0.01779584027826786
step: 110, loss: 0.006236860994249582
step: 120, loss: 9.134155698120594e-05
step: 130, loss: 5.864181730430573e-05
step: 140, loss: 9.961133764591068e-05
step: 150, loss: 0.0005964243318885565
step: 160, loss: 3.159698098897934e-05
step: 170, loss: 7.62153067626059e-05
step: 180, loss: 6.597331230295822e-05
step: 190, loss: 0.0010515146423131227
step: 200, loss: 0.0016304283635690808
step: 210, loss: 0.021363217383623123
step: 220, loss: 7.494154124287888e-05
step: 230, loss: 2.8672884582192637e-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=0.9810055865921787, f1=0.9733333333333333,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007267416920512915
step: 10, loss: 0.000966420047916472
step: 20, loss: 4.8667341616237536e-05
step: 30, loss: 0.001029093051329255
step: 40, loss: 5.275895819067955e-05
step: 50, loss: 4.662742867367342e-05
step: 60, loss: 0.0013674342771992087
step: 70, loss: 0.00032020072103478014
step: 80, loss: 0.04428461194038391
step: 90, loss: 2.8911199478898197e-05
step: 100, loss: 7.53225467633456e-05

```

```

step: 110, loss: 7.332704990403727e-05
step: 120, loss: 3.9430302422260866e-05
step: 130, loss: 4.1889325075317174e-05
step: 140, loss: 0.043120622634887695
step: 150, loss: 6.197665061336011e-05
step: 160, loss: 0.0007786840433254838
step: 170, loss: 3.663317329483107e-05
step: 180, loss: 5.7333119912073016e-05
step: 190, loss: 0.010310468263924122
step: 200, loss: 8.199531293939799e-05
step: 210, loss: 0.0010078795021399856
step: 220, loss: 0.018460772931575775
step: 230, loss: 0.00010401221516076475
/usr/local/lib/python3.7/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=0.9799554565701558, f1=0.9722530521642618,
best_f1=0.9700332963374029
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 [00:08, 289.92it/s]
load_f1 = 0.9810479375696767
real_f1 = 0.9799554565701558
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 [00:12, 344.51it/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 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: 0.7902925610542297
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.464571088552475
step: 20, loss: 0.49128106236457825
step: 30, loss: 0.3258640468120575
step: 40, loss: 0.27026861906051636
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.19615904986858368
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 4096.0
step: 60, loss: 0.36191239953041077
step: 70, loss: 0.12090165168046951
step: 80, loss: 0.13441914319992065
step: 90, loss: 0.14243651926517487
step: 100, loss: 0.3202132284641266

```

```
step: 110, loss: 0.022112248465418816
step: 120, loss: 0.046944405883550644
step: 130, loss: 0.011157985776662827
step: 140, loss: 0.22613203525543213
step: 150, loss: 0.022609852254390717
step: 160, loss: 0.15006208419799805
step: 170, loss: 0.14695094525814056
step: 180, loss: 0.09882020205259323
step: 190, loss: 0.02557608298957348
step: 200, loss: 0.07720357924699783
step: 210, loss: 0.0828058272600174
step: 220, loss: 0.09844157844781876
step: 230, loss: 0.13218143582344055
step: 240, loss: 0.09093205630779266
step: 250, loss: 0.07007547467947006
step: 260, loss: 0.038361113518476486
step: 270, loss: 0.01093970611691475
step: 280, loss: 0.08165781944990158
step: 290, loss: 0.03361085057258606
step: 300, loss: 0.20363067090511322
step: 310, loss: 0.07700033485889435
step: 320, loss: 0.06428172439336777
step: 330, loss: 0.13029690086841583
step: 340, loss: 0.12140527367591858
step: 350, loss: 0.11739061772823334
step: 360, loss: 0.08404663950204849
step: 370, loss: 0.20197153091430664
step: 380, loss: 0.17292992770671844
step: 390, loss: 0.009978339076042175
step: 400, loss: 0.011762365698814392
step: 410, loss: 0.01305331476032734
step: 420, loss: 0.014612465165555477
step: 430, loss: 0.10111576318740845
step: 440, loss: 0.052113935351371765
step: 450, loss: 0.022560998797416687
step: 460, loss: 0.19869522750377655
step: 470, loss: 0.22887183725833893
step: 480, loss: 0.27735182642936707
step: 490, loss: 0.03899084031581879
step: 500, loss: 0.0068472824059426785
step: 510, loss: 0.022817017510533333
step: 520, loss: 0.04394877329468727
step: 530, loss: 0.06263522058725357
/usr/local/lib/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=0.938134810710988, f1=0.9383974062065772,
best_f1=0.9383974062065772
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15215322375297546
step: 10, loss: 0.34124234318733215
step: 20, loss: 0.037004504352808
step: 30, loss: 0.0342056043446064
step: 40, loss: 0.012230568565428257
step: 50, loss: 0.09675686806440353
step: 60, loss: 0.055535007268190384
step: 70, loss: 0.08873105049133301
step: 80, loss: 0.021469438448548317
step: 90, loss: 0.004578379448503256
step: 100, loss: 0.2791077494621277
step: 110, loss: 0.034463513642549515
step: 120, loss: 0.17062295973300934
step: 130, loss: 0.036240704357624054
step: 140, loss: 0.021411066874861717
step: 150, loss: 0.12411507219076157
step: 160, loss: 0.01970740593969822
step: 170, loss: 0.039743248373270035
step: 180, loss: 0.003924158867448568
step: 190, loss: 0.007999724708497524
step: 200, loss: 0.048121046274900436
step: 210, loss: 0.004974985495209694
step: 220, loss: 0.14898346364498138
step: 230, loss: 0.009719938971102238
step: 240, loss: 0.10920150578022003
step: 250, loss: 0.05360572412610054
step: 260, loss: 0.08101608604192734
step: 270, loss: 0.14079119265079498
step: 280, loss: 0.12681949138641357
step: 290, loss: 0.03471832349896431
step: 300, loss: 0.03319493681192398
step: 310, loss: 0.025028999894857407
step: 320, loss: 0.06219679117202759
step: 330, loss: 0.04547002166509628
step: 340, loss: 0.013692328706383705
step: 350, loss: 0.030234500765800476
step: 360, loss: 0.01892276480793953
step: 370, loss: 0.00562105281278491
step: 380, loss: 0.07584510743618011
step: 390, loss: 0.0890098437666893
step: 400, loss: 0.022571612149477005

```

```

step: 410, loss: 0.0016465387307107449
step: 420, loss: 0.031370870769023895
step: 430, loss: 0.03613658621907234
step: 440, loss: 0.00856068730354309
step: 450, loss: 0.08329629898071289
step: 460, loss: 0.2629527151584625
step: 470, loss: 0.14103823900222778
step: 480, loss: 0.17161640524864197
step: 490, loss: 0.025271670892834663
step: 500, loss: 0.07086239755153656
step: 510, loss: 0.07904447615146637
step: 520, loss: 0.12510761618614197
step: 530, loss: 0.11430075019598007
/usr/local/lib/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=0.9465930018416207, f1=0.9463548830811555,
best_f1=0.9463548830811555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009830170311033726
step: 10, loss: 0.03781736642122269
step: 20, loss: 0.11594990640878677
step: 30, loss: 0.30480802059173584
step: 40, loss: 0.015931937843561172
step: 50, loss: 0.010361386463046074
step: 60, loss: 0.016881754621863365
step: 70, loss: 0.11554085463285446
step: 80, loss: 0.0043894024565815926
step: 90, loss: 0.11168655008077621
step: 100, loss: 0.013847279362380505
step: 110, loss: 0.03722911700606346
step: 120, loss: 0.006949760019779205
step: 130, loss: 0.010543884709477425
step: 140, loss: 0.013096398673951626
step: 150, loss: 0.027617132291197777
step: 160, loss: 0.010809305123984814
step: 170, loss: 0.055749595165252686
step: 180, loss: 0.021163521334528923
step: 190, loss: 0.005567984189838171
step: 200, loss: 0.018740618601441383
step: 210, loss: 0.026737840846180916
step: 220, loss: 0.027723899111151695

```

```
step: 230, loss: 0.01618906296789646
step: 240, loss: 0.045384958386421204
step: 250, loss: 0.003540421836078167
step: 260, loss: 0.01689930632710457
step: 270, loss: 0.004562088288366795
step: 280, loss: 0.005091926548629999
step: 290, loss: 0.05941120535135269
step: 300, loss: 0.1791858971118927
step: 310, loss: 0.21785211563110352
step: 320, loss: 0.18009920418262482
step: 330, loss: 0.0021172582637518644
step: 340, loss: 0.003514935029670596
step: 350, loss: 0.03261163830757141
step: 360, loss: 0.005945135373622179
step: 370, loss: 0.0025558150373399258
step: 380, loss: 0.07146311551332474
step: 390, loss: 0.006339884828776121
step: 400, loss: 0.03584279119968414
step: 410, loss: 0.009242166765034199
step: 420, loss: 0.08716893196105957
step: 430, loss: 0.06312499195337296
step: 440, loss: 0.02293127216398716
step: 450, loss: 0.04739968851208687
step: 460, loss: 0.1936110109090805
step: 470, loss: 0.005287709180265665
step: 480, loss: 0.012817509472370148
step: 490, loss: 0.009851557202637196
step: 500, loss: 0.2129720151424408
step: 510, loss: 0.011320237070322037
step: 520, loss: 0.01250970084220171
step: 530, loss: 0.011125179007649422
/usr/local/lib/python3.7/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=0.9442870632672333, f1=0.9395656279508972,
best_f1=0.9463548830811555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021988668013364077
step: 10, loss: 0.005342434160411358
step: 20, loss: 0.14378635585308075
step: 30, loss: 0.0012890780344605446
step: 40, loss: 0.0013805069029331207
step: 50, loss: 0.0025291091296821833
step: 60, loss: 0.0019150448497384787
step: 70, loss: 0.004520073067396879
step: 80, loss: 0.018819451332092285
```



```
step: 90, loss: 0.13225096464157104
step: 100, loss: 0.011330788023769855
step: 110, loss: 0.017696108669042587
step: 120, loss: 0.004392606671899557
step: 130, loss: 0.003205915680155158
step: 140, loss: 0.017246397212147713
step: 150, loss: 0.0022527577821165323
step: 160, loss: 0.042580828070640564
step: 170, loss: 0.011688115075230598
step: 180, loss: 0.007113915868103504
step: 190, loss: 0.013136876747012138
step: 200, loss: 0.002329102251678705
step: 210, loss: 0.09490928053855896
step: 220, loss: 0.006626965943723917
step: 230, loss: 0.18463075160980225
step: 240, loss: 0.016168100759387016
step: 250, loss: 0.0028046160005033016
step: 260, loss: 0.10802341997623444
step: 270, loss: 0.0058420076966285706
step: 280, loss: 0.0036189656239002943
step: 290, loss: 0.08403937518596649
step: 300, loss: 0.01191391795873642
step: 310, loss: 0.08743510395288467
step: 320, loss: 0.16232535243034363
step: 330, loss: 0.044390350580215454
step: 340, loss: 0.008570666424930096
step: 350, loss: 0.05599845573306084
step: 360, loss: 0.008924761787056923
step: 370, loss: 0.036845121532678604
step: 380, loss: 0.005285118240863085
step: 390, loss: 0.11689461022615433
step: 400, loss: 0.0050492300651967525
step: 410, loss: 0.0586824007332325
step: 420, loss: 0.0047194804064929485
step: 430, loss: 0.00649133650586009
step: 440, loss: 0.050774384289979935
step: 450, loss: 0.014608463272452354
step: 460, loss: 0.013653167523443699
step: 470, loss: 0.02760719694197178
step: 480, loss: 0.003965894691646099
step: 490, loss: 0.04953759163618088
step: 500, loss: 0.019969098269939423
step: 510, loss: 0.12847770750522614
step: 520, loss: 0.0052710226736962795
step: 530, loss: 0.002419473836198449
/usr/local/lib/python3.7/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=0.9399161620866325, f1=0.9414498141263941,  
best_f1=0.9463548830811555  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.003750499803572893  
step: 10, loss: 0.0020672017708420753  
step: 20, loss: 0.004267506767064333  
step: 30, loss: 0.0028219332452863455  
step: 40, loss: 0.10958646237850189  
step: 50, loss: 0.008326847106218338  
step: 60, loss: 0.005529599729925394  
step: 70, loss: 0.0006461184821091592  
step: 80, loss: 0.00721705611795187  
step: 90, loss: 0.0020586929749697447  
step: 100, loss: 0.09073040634393692  
step: 110, loss: 0.0016709071351215243  
step: 120, loss: 0.014982318505644798  
step: 130, loss: 0.0029724803753197193  
step: 140, loss: 0.0029061369132250547  
step: 150, loss: 0.003253207542002201  
step: 160, loss: 0.022628189995884895  
step: 170, loss: 0.016701651737093925  
step: 180, loss: 0.003136534709483385  
step: 190, loss: 0.0002937925746664405  
step: 200, loss: 0.0006899531581439078  
step: 210, loss: 0.04491525888442993  
step: 220, loss: 0.0007903505465947092  
step: 230, loss: 0.0022894288413226604  
step: 240, loss: 0.024345586076378822  
step: 250, loss: 0.0007424478535540402  
step: 260, loss: 0.05763882026076317  
step: 270, loss: 0.00565605191513896  
step: 280, loss: 0.0641293153166771  
step: 290, loss: 0.0012840129202231765  
step: 300, loss: 0.04025533050298691  
step: 310, loss: 0.002135137328878045  
step: 320, loss: 0.0007855836884118617  
step: 330, loss: 0.0035752742551267147  
step: 340, loss: 0.00482152821496129  
step: 350, loss: 0.014462661929428577  
step: 360, loss: 0.0021227411925792694  
step: 370, loss: 0.0022834395058453083  
step: 380, loss: 0.07004215568304062  
step: 390, loss: 0.0012458740966394544  
step: 400, loss: 0.01296248659491539  
step: 410, loss: 0.00041300265002064407  
step: 420, loss: 0.0030425209552049637
```

```
step: 430, loss: 0.009102311916649342
step: 440, loss: 0.0018081639427691698
step: 450, loss: 0.035069774836301804
step: 460, loss: 0.013938595540821552
step: 470, loss: 0.05243684723973274
step: 480, loss: 0.021124232560396194
step: 490, loss: 0.03556757792830467
step: 500, loss: 0.027153052389621735
step: 510, loss: 0.009706876240670681
step: 520, loss: 0.0005019917152822018
step: 530, loss: 0.002608662936836481
/usr/local/lib/python3.7/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=0.9442100328176277, f1=0.9435559736594544,
best_f1=0.9463548830811555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003287800820544362
step: 10, loss: 0.00297866971231997
step: 20, loss: 0.002885288093239069
step: 30, loss: 0.007278843317180872
step: 40, loss: 0.013649463653564453
step: 50, loss: 0.004631364718079567
step: 60, loss: 0.004658065736293793
step: 70, loss: 0.009694833308458328
step: 80, loss: 0.0007212257478386164
step: 90, loss: 0.0006033005192875862
step: 100, loss: 0.00396004831418395
step: 110, loss: 0.007857531309127808
step: 120, loss: 0.0019983036909252405
step: 130, loss: 0.0011672989930957556
step: 140, loss: 0.004426895175129175
step: 150, loss: 0.005954817868769169
step: 160, loss: 0.0001315035915467888
step: 170, loss: 0.004983106162399054
step: 180, loss: 0.00029412980074994266
step: 190, loss: 0.00512762600556016
step: 200, loss: 0.00039892681525088847
step: 210, loss: 0.002721728291362524
step: 220, loss: 0.00184251694008708
step: 230, loss: 0.000967249448876828
step: 240, loss: 0.02066025696694851
step: 250, loss: 0.037412840873003006
step: 260, loss: 0.001773766358383
step: 270, loss: 0.0045464495196938515
step: 280, loss: 0.002072955947369337
```

```

step: 290, loss: 0.0003853872185572982
step: 300, loss: 0.00032293808180838823
step: 310, loss: 0.0005468994495458901
step: 320, loss: 0.011068332009017467
step: 330, loss: 0.0024456714745610952
step: 340, loss: 0.0222407728433609
step: 350, loss: 0.020575525239109993
step: 360, loss: 0.02493675984442234
step: 370, loss: 0.0004800432943738997
step: 380, loss: 0.006437754258513451
step: 390, loss: 0.0007709775236435235
step: 400, loss: 0.0022389821242541075
step: 410, loss: 0.006814983673393726
step: 420, loss: 0.0005222067120485008
step: 430, loss: 0.00043679718510247767
step: 440, loss: 0.014512213878333569
step: 450, loss: 0.0005713429418392479
step: 460, loss: 0.01702284812927246
step: 470, loss: 0.07757624983787537
step: 480, loss: 0.012389583513140678
step: 490, loss: 0.0004448713443707675
step: 500, loss: 0.004691971931606531
step: 510, loss: 0.0002233613922726363
step: 520, loss: 0.020434318110346794
step: 530, loss: 0.0157791655510664
/usr/local/lib/python3.7/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=0.9387186629526463, f1=0.9386248269497001,
best_f1=0.9463548830811555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019249921897426248
step: 10, loss: 0.035921674221754074
step: 20, loss: 0.014268890023231506
step: 30, loss: 0.021502962335944176
step: 40, loss: 0.0014445367269217968
step: 50, loss: 0.002190880011767149
step: 60, loss: 0.0035253504756838083
step: 70, loss: 0.06858621537685394
step: 80, loss: 0.17341803014278412
step: 90, loss: 0.0014114242512732744
step: 100, loss: 0.0139748714864254
step: 110, loss: 0.015615090727806091
step: 120, loss: 0.0006816974491812289
step: 130, loss: 0.0017526502488180995
step: 140, loss: 0.007909269072115421

```

```
step: 150, loss: 0.00014009192818775773
step: 160, loss: 0.006132378242909908
step: 170, loss: 0.0038965505082160234
step: 180, loss: 0.0002641117316670716
step: 190, loss: 0.005780784413218498
step: 200, loss: 0.004524163901805878
step: 210, loss: 0.0037570660933852196
step: 220, loss: 0.04708639159798622
step: 230, loss: 0.001172534073702991
step: 240, loss: 0.003204131266102195
step: 250, loss: 0.036991193890571594
step: 260, loss: 0.0002457946538925171
step: 270, loss: 0.00047245173482224345
step: 280, loss: 0.07340866327285767
step: 290, loss: 0.003749858122318983
step: 300, loss: 0.00028002611361443996
step: 310, loss: 0.0011903245467692614
step: 320, loss: 0.04649822786450386
step: 330, loss: 0.008110960945487022
step: 340, loss: 0.00466365460306406
step: 350, loss: 0.00043125866795890033
step: 360, loss: 0.0037686466239392757
step: 370, loss: 0.0037878293078392744
step: 380, loss: 0.0044225468300282955
step: 390, loss: 0.0015923265600576997
step: 400, loss: 0.00023175695969257504
step: 410, loss: 0.0011347285471856594
step: 420, loss: 0.0001936078624567017
step: 430, loss: 0.0007574825431220233
step: 440, loss: 0.00299329636618495
step: 450, loss: 0.014361520297825336
step: 460, loss: 0.01661922037601471
step: 470, loss: 0.05158122256398201
step: 480, loss: 0.0003716540231835097
step: 490, loss: 0.0011710576945915818
step: 500, loss: 8.548675396014005e-05
step: 510, loss: 0.0007412440609186888
step: 520, loss: 0.0011354117887094617
step: 530, loss: 0.0002059941180050373
/usr/local/lib/python3.7/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=0.9446009389671362, f1=0.9395784543325526,
best_f1=0.9463548830811555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018693192396312952
```

step: 10, loss: 0.0044958321377635
step: 20, loss: 0.01976616121828556
step: 30, loss: 0.05219843611121178
step: 40, loss: 0.0007756247068755329
step: 50, loss: 0.00015342146798502654
step: 60, loss: 0.0009110654937103391
step: 70, loss: 0.0039065624587237835
step: 80, loss: 0.0003810189664363861
step: 90, loss: 0.00018414044461678714
step: 100, loss: 0.015454926528036594
step: 110, loss: 0.0006157358875498176
step: 120, loss: 0.014737972058355808
step: 130, loss: 0.03285776451230049
step: 140, loss: 8.15774328657426e-05
step: 150, loss: 0.001068290090188384
step: 160, loss: 0.0018397177336737514
step: 170, loss: 0.13372264802455902
step: 180, loss: 0.0038194574881345034
step: 190, loss: 0.003540300764143467
step: 200, loss: 0.0020746104419231415
step: 210, loss: 0.0007524157990701497
step: 220, loss: 0.0005859275697730482
step: 230, loss: 0.00815651472657919
step: 240, loss: 0.00028262799605727196
step: 250, loss: 0.04180677607655525
step: 260, loss: 0.22317834198474884
step: 270, loss: 0.000368524924851954
step: 280, loss: 0.0018887650221586227
step: 290, loss: 0.011991147883236408
step: 300, loss: 0.0005421806708909571
step: 310, loss: 0.028991110622882843
step: 320, loss: 0.0024179627653211355
step: 330, loss: 0.01660725846886635
step: 340, loss: 0.0026908174622803926
step: 350, loss: 0.024382147938013077
step: 360, loss: 0.001489858259446919
step: 370, loss: 0.0025310202036052942
step: 380, loss: 0.009254446253180504
step: 390, loss: 0.000766711076721549
step: 400, loss: 0.15080997347831726
step: 410, loss: 0.00034533266443759203
step: 420, loss: 0.00014664078480564058
step: 430, loss: 9.884398605208844e-05
step: 440, loss: 0.008142014034092426
step: 450, loss: 0.0032786070369184017
step: 460, loss: 0.003753806697204709
step: 470, loss: 0.00028541882056742907
step: 480, loss: 0.00046310812467709184

```

step: 490, loss: 0.00016976242477539927
step: 500, loss: 0.004990996792912483
step: 510, loss: 0.003317989408969879
step: 520, loss: 0.0004167855076957494
step: 530, loss: 9.998838504543528e-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 8: dev_f1=0.9489795918367347, f1=0.9391143911439114,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011389061546651646
step: 10, loss: 0.00040355740929953754
step: 20, loss: 0.0002644876658450812
step: 30, loss: 0.00021747926075477153
step: 40, loss: 0.002372013172134757
step: 50, loss: 0.0003385973395779729
step: 60, loss: 0.0005271655973047018
step: 70, loss: 0.11498428881168365
step: 80, loss: 0.00027861815760843456
step: 90, loss: 0.0006618132465519011
step: 100, loss: 0.00012249842984601855
step: 110, loss: 0.000916916411370039
step: 120, loss: 0.0005769857089035213
step: 130, loss: 0.003408597083762288
step: 140, loss: 0.061963435262441635
step: 150, loss: 0.0006068520015105605
step: 160, loss: 0.01374791469424963
step: 170, loss: 0.0012055858969688416
step: 180, loss: 0.00014158901467453688
step: 190, loss: 0.019278259947896004
step: 200, loss: 0.00023190997308120131
step: 210, loss: 0.0009756487561389804
step: 220, loss: 0.0005221145693212748
step: 230, loss: 8.70375006343238e-05
step: 240, loss: 0.0001405471411999315
step: 250, loss: 0.0019372687675058842
step: 260, loss: 0.07826859503984451
step: 270, loss: 0.0035826952662318945
step: 280, loss: 0.00013278140977490693
step: 290, loss: 0.0004033804580103606
step: 300, loss: 0.0003437318082433194

```

```

step: 310, loss: 0.00017456745263189077
step: 320, loss: 0.0018557546427473426
step: 330, loss: 0.0013382823672145605
step: 340, loss: 0.000788295641541481
step: 350, loss: 0.0028002706822007895
step: 360, loss: 0.04383543133735657
step: 370, loss: 0.0003887044149450958
step: 380, loss: 0.0001305088953813538
step: 390, loss: 0.00035541196120902896
step: 400, loss: 0.0016415058635175228
step: 410, loss: 0.006961253937333822
step: 420, loss: 0.00014722251216880977
step: 430, loss: 0.00022270642512012273
step: 440, loss: 0.0004436706658452749
step: 450, loss: 7.392761472146958e-05
step: 460, loss: 0.0005348570994101465
step: 470, loss: 5.299284021020867e-05
step: 480, loss: 6.154493894428015e-05
step: 490, loss: 0.0004657392273657024
step: 500, loss: 0.06243368983268738
step: 510, loss: 0.0004638218379113823
step: 520, loss: 0.00010591973841656
step: 530, loss: 0.001430604374036193
/usr/local/lib/python3.7/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=0.9457579972183587, f1=0.942095588235294,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00167948380112648
step: 10, loss: 0.0002739481860771775
step: 20, loss: 0.00012308800069149584
step: 30, loss: 0.00012230964784976095
step: 40, loss: 7.357799040619284e-05
step: 50, loss: 0.0003130270924884826
step: 60, loss: 0.00033096253173425794
step: 70, loss: 0.0004890643176622689
step: 80, loss: 0.0007475868915207684
step: 90, loss: 0.0002901373663917184
step: 100, loss: 0.0004358227306511253
step: 110, loss: 0.000388407614082098
step: 120, loss: 0.00011762202484533191
step: 130, loss: 0.0001373265840811655
step: 140, loss: 0.0007631418411619961
step: 150, loss: 0.0001308221253566444
step: 160, loss: 5.0812748668249696e-05

```



```

step: 170, loss: 9.068863437278196e-05
step: 180, loss: 0.014267458580434322
step: 190, loss: 0.0019106828840449452
step: 200, loss: 0.0001412331621395424
step: 210, loss: 0.0016731637297198176
step: 220, loss: 9.136803419096395e-05
step: 230, loss: 6.900472362758592e-05
step: 240, loss: 9.891628724290058e-05
step: 250, loss: 0.0017306109657511115
step: 260, loss: 0.0009148882236331701
step: 270, loss: 0.00021126704814378172
step: 280, loss: 2.849014344974421e-05
step: 290, loss: 8.452519978163764e-05
step: 300, loss: 0.00039545222534798086
step: 310, loss: 0.00024654497974552214
step: 320, loss: 0.0007739875582046807
step: 330, loss: 0.007882599718868732
step: 340, loss: 0.008171981200575829
step: 350, loss: 0.000399122858652845
step: 360, loss: 0.0029006891418248415
step: 370, loss: 0.0010663170833140612
step: 380, loss: 0.0004211501218378544
step: 390, loss: 0.0006058905855752528
step: 400, loss: 0.00048692108248360455
step: 410, loss: 0.01583947241306305
step: 420, loss: 0.0063411761075258255
step: 430, loss: 0.00019580671505536884
step: 440, loss: 0.00047643075231462717
step: 450, loss: 0.007371263112872839
step: 460, loss: 0.002709226217120886
step: 470, loss: 0.0031781778670847416
step: 480, loss: 0.048245515674352646
step: 490, loss: 0.12812866270542145
step: 500, loss: 0.10645288974046707
step: 510, loss: 0.0005360232316888869
step: 520, loss: 0.0005438461666926742
step: 530, loss: 0.0011376640759408474
/usr/local/lib/python3.7/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=0.9472718936267768, f1=0.9433106575963719,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005456707440316677
step: 10, loss: 0.05292786285281181
step: 20, loss: 0.0003563560894690454

```

step: 30, loss: 0.004083351697772741
step: 40, loss: 0.01218595914542675
step: 50, loss: 0.00031163758831098676
step: 60, loss: 0.0003160855558235198
step: 70, loss: 6.772441702196375e-05
step: 80, loss: 0.0028250031173229218
step: 90, loss: 0.0023170732893049717
step: 100, loss: 0.000558609957806766
step: 110, loss: 0.00010466855019330978
step: 120, loss: 0.006359125021845102
step: 130, loss: 0.0008278420427814126
step: 140, loss: 0.0001614179345779121
step: 150, loss: 0.0015889675123617053
step: 160, loss: 0.00034813175443559885
step: 170, loss: 0.048567984253168106
step: 180, loss: 7.060031202854589e-05
step: 190, loss: 0.006865170784294605
step: 200, loss: 0.002547411946579814
step: 210, loss: 5.996382242301479e-05
step: 220, loss: 0.0002949991903733462
step: 230, loss: 0.0025203723926097155
step: 240, loss: 6.716627831337973e-05
step: 250, loss: 6.153163849376142e-05
step: 260, loss: 3.7638412322849035e-05
step: 270, loss: 0.0007589157903566957
step: 280, loss: 0.000464037642814219
step: 290, loss: 0.007503077387809753
step: 300, loss: 0.0037466983776539564
step: 310, loss: 3.8129590393509716e-05
step: 320, loss: 0.02290869876742363
step: 330, loss: 0.0004977795761078596
step: 340, loss: 0.0016123456880450249
step: 350, loss: 0.0013898792676627636
step: 360, loss: 0.001490659429691732
step: 370, loss: 4.644695945899002e-05
step: 380, loss: 0.00016728938498999923
step: 390, loss: 0.012189436703920364
step: 400, loss: 9.53356793615967e-05
step: 410, loss: 0.0015877485275268555
step: 420, loss: 0.007387079764157534
step: 430, loss: 0.00039299720083363354
step: 440, loss: 0.0007309510256163776
step: 450, loss: 0.0003557992458809167
step: 460, loss: 0.01020633615553379
step: 470, loss: 0.0005391298327594995
step: 480, loss: 0.004094969481229782
step: 490, loss: 0.002623308217152953
step: 500, loss: 0.00019129201245959848

```

step: 510, loss: 8.97631689440459e-05
step: 520, loss: 4.385616193758324e-05
step: 530, loss: 5.396423512138426e-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=0.943609022556391, f1=0.9350770667912189,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001522599719464779
step: 10, loss: 6.562735507031903e-05
step: 20, loss: 3.859343269141391e-05
step: 30, loss: 0.00042155335540883243
step: 40, loss: 0.00012619821063708514
step: 50, loss: 0.0006270254962146282
step: 60, loss: 3.573679714463651e-05
step: 70, loss: 0.00014033888874109834
step: 80, loss: 0.0019581408705562353
step: 90, loss: 0.004410210065543652
step: 100, loss: 0.0017888686852529645
step: 110, loss: 2.723489342315588e-05
step: 120, loss: 0.0004976126365363598
step: 130, loss: 0.002325458452105522
step: 140, loss: 0.00016733797383494675
step: 150, loss: 0.0004327250353526324
step: 160, loss: 6.41035585431382e-05
step: 170, loss: 0.0001631969353184104
step: 180, loss: 0.0003056832938455045
step: 190, loss: 0.008963568136096
step: 200, loss: 0.00013224191206973046
step: 210, loss: 0.0013710001949220896
step: 220, loss: 4.067247937200591e-05
step: 230, loss: 0.003121806774288416
step: 240, loss: 0.0006167361862026155
step: 250, loss: 0.0004185567086096853
step: 260, loss: 0.000794800347648561
step: 270, loss: 0.00045064702862873673
step: 280, loss: 0.0005360639770515263
step: 290, loss: 0.0008118526893667877
step: 300, loss: 0.003483908250927925
step: 310, loss: 0.00010471901623532176
step: 320, loss: 0.0002558867272455245
step: 330, loss: 0.0023865248076617718
step: 340, loss: 0.0020051635801792145
step: 350, loss: 0.01181703619658947
step: 360, loss: 0.0009739025263115764

```

```

step: 370, loss: 9.322829282609746e-05
step: 380, loss: 0.0011751544661819935
step: 390, loss: 0.00014321996422950178
step: 400, loss: 0.0003617359325289726
step: 410, loss: 0.0076791369356215
step: 420, loss: 0.0037910682149231434
step: 430, loss: 0.00017074756033252925
step: 440, loss: 0.008908684365451336
step: 450, loss: 0.0053122020326554775
step: 460, loss: 0.0059096976183354855
step: 470, loss: 7.801967876730487e-05
step: 480, loss: 0.0019722117576748133
step: 490, loss: 0.0017505655996501446
step: 500, loss: 0.0048867519944906235
step: 510, loss: 0.00018433693912811577
step: 520, loss: 0.01782563328742981
step: 530, loss: 0.0009347500745207071
/usr/local/lib/python3.7/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=0.9465020576131687, f1=0.9405356332274172,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004105093830730766
step: 10, loss: 0.0005574798560701311
step: 20, loss: 0.0481172539293766
step: 30, loss: 0.0017994707450270653
step: 40, loss: 0.0007350933155976236
step: 50, loss: 0.0008977752877399325
step: 60, loss: 0.00033474553492851555
step: 70, loss: 0.00013548799324780703
step: 80, loss: 0.0018195307347923517
step: 90, loss: 0.0022115635219961405
step: 100, loss: 6.339103856589645e-05
step: 110, loss: 0.0005946239689365029
step: 120, loss: 0.00016062657232396305
step: 130, loss: 0.002529725432395935
step: 140, loss: 5.452868572319858e-05
step: 150, loss: 0.011389711871743202
step: 160, loss: 0.02138812281191349
step: 170, loss: 6.537316949106753e-05
step: 180, loss: 0.00041298658470623195
step: 190, loss: 0.00017205336189363152
step: 200, loss: 0.0027117948047816753
step: 210, loss: 0.0023244735784828663
step: 220, loss: 0.0007559827645309269

```

```

step: 230, loss: 0.007542045786976814
step: 240, loss: 2.995605427713599e-05
step: 250, loss: 0.02891518548130989
step: 260, loss: 0.0016458774916827679
step: 270, loss: 0.00040961295599117875
step: 280, loss: 0.003816361539065838
step: 290, loss: 0.00013747485354542732
step: 300, loss: 0.0005863456171937287
step: 310, loss: 0.0003204231325071305
step: 320, loss: 2.0481391402427107e-05
step: 330, loss: 0.00198126258328557
step: 340, loss: 2.6925670681521297e-05
step: 350, loss: 0.23394757509231567
step: 360, loss: 5.072347994428128e-05
step: 370, loss: 0.00011215870472369716
step: 380, loss: 0.0016172417672351003
step: 390, loss: 0.00016306077304761857
step: 400, loss: 0.0005601823795586824
step: 410, loss: 0.00010886683594435453
step: 420, loss: 0.021356604993343353
step: 430, loss: 0.0015409872867166996
step: 440, loss: 0.00021615774312522262
step: 450, loss: 0.0020142041612416506
step: 460, loss: 4.534739491646178e-05
step: 470, loss: 0.01723170280456543
step: 480, loss: 0.0024016748648136854
step: 490, loss: 0.0002432005712762475
step: 500, loss: 0.00018065473705064505
step: 510, loss: 5.427721771411598e-05
step: 520, loss: 4.588416777551174e-05
step: 530, loss: 0.00023493827029597014
/usr/local/lib/python3.7/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=0.945736434108527, f1=0.9402173913043477,
best_f1=0.9391143911439114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023730202519800514
step: 10, loss: 0.003479116829112172
step: 20, loss: 7.602482219226658e-05
step: 30, loss: 0.0006768520106561482
step: 40, loss: 0.0026961080729961395
step: 50, loss: 0.0012919916771352291
step: 60, loss: 8.868938311934471e-05
step: 70, loss: 0.00018668099073693156
step: 80, loss: 0.00929130706936121

```

```
step: 90, loss: 0.00011988924234174192
step: 100, loss: 0.00017064352869056165
step: 110, loss: 0.0002005725837079808
step: 120, loss: 0.00019950693240389228
step: 130, loss: 0.003901182673871517
step: 140, loss: 0.0015350556932389736
step: 150, loss: 0.009452532976865768
step: 160, loss: 0.0020242466125637293
step: 170, loss: 0.0006133801070973277
step: 180, loss: 0.00037303270073607564
step: 190, loss: 0.0014170040376484394
step: 200, loss: 0.010375909507274628
step: 210, loss: 0.0009207176044583321
step: 220, loss: 0.11954428255558014
step: 230, loss: 0.0017189406789839268
step: 240, loss: 0.004467739723622799
step: 250, loss: 8.3904255006928e-05
step: 260, loss: 7.929633284220472e-05
step: 270, loss: 0.004073238465934992
step: 280, loss: 0.00027757344651035964
step: 290, loss: 0.00020470631716307253
step: 300, loss: 0.009874795563519001
step: 310, loss: 4.7522767999907956e-05
step: 320, loss: 0.011607381515204906
step: 330, loss: 0.0003588135587051511
step: 340, loss: 4.20399010181427e-05
step: 350, loss: 0.0011907462030649185
step: 360, loss: 0.0012519031297415495
step: 370, loss: 0.00020634553220588714
step: 380, loss: 0.003009797539561987
step: 390, loss: 0.00016653178317938
step: 400, loss: 3.541787373251282e-05
step: 410, loss: 0.0001562807010486722
step: 420, loss: 0.00016145585686899722
step: 430, loss: 0.0003353656211402267
step: 440, loss: 8.598098065704107e-05
step: 450, loss: 0.00019540690118446946
step: 460, loss: 0.01329416036605835
step: 470, loss: 3.268799264333211e-05
step: 480, loss: 8.370433351956308e-05
step: 490, loss: 0.00033069855999201536
step: 500, loss: 9.239554492523894e-05
step: 510, loss: 8.725956286070868e-05
step: 520, loss: 2.198238689743448e-05
step: 530, loss: 4.018185427412391e-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=0.9473193473193473, f1=0.9405756731662025,  
best_f1=0.9391143911439114  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 8.974804222816601e-05  
step: 10, loss: 2.3807613615645096e-05  
step: 20, loss: 0.00015955694834701717  
step: 30, loss: 0.0012829642510041595  
step: 40, loss: 4.725222606793977e-05  
step: 50, loss: 0.0004137979994993657  
step: 60, loss: 2.8121563445893116e-05  
step: 70, loss: 0.0006768692983314395  
step: 80, loss: 4.607069422490895e-05  
step: 90, loss: 1.8171755073126405e-05  
step: 100, loss: 7.086939149303362e-05  
step: 110, loss: 0.009921674616634846  
step: 120, loss: 0.0017506062285974622  
step: 130, loss: 3.435234248172492e-05  
step: 140, loss: 4.2333242163294926e-05  
step: 150, loss: 0.00022536673350259662  
step: 160, loss: 4.5095846871845424e-05  
step: 170, loss: 5.089402839075774e-05  
step: 180, loss: 0.0017667480278760195  
step: 190, loss: 0.00018701939552556723  
step: 200, loss: 0.00015742414689157158  
step: 210, loss: 0.0001583928387844935  
step: 220, loss: 0.00011722560157068074  
step: 230, loss: 8.711377449799329e-05  
step: 240, loss: 0.00014637864660471678  
step: 250, loss: 7.240489503601566e-05  
step: 260, loss: 0.022458210587501526  
step: 270, loss: 0.0014334406005218625  
step: 280, loss: 0.0012249695137143135  
step: 290, loss: 0.0018792058108374476  
step: 300, loss: 0.00014383127563633025  
step: 310, loss: 0.0004405398794915527  
step: 320, loss: 0.00016419723397120833  
step: 330, loss: 0.00014882431423757225  
step: 340, loss: 0.0002758628106676042  
step: 350, loss: 0.00029969721799716353  
step: 360, loss: 0.00022520664788316935  
step: 370, loss: 4.566321149468422e-05  
step: 380, loss: 0.00017945239960681647  
step: 390, loss: 0.030529603362083435  
step: 400, loss: 0.003133717691525817  
step: 410, loss: 0.0006183492369018495  
step: 420, loss: 6.295352068264037e-05
```

```

step: 430, loss: 0.0001856636517914012
step: 440, loss: 0.0005905400030314922
step: 450, loss: 0.0036784205585718155
step: 460, loss: 9.678786591393873e-05
step: 470, loss: 0.0008033909252844751
step: 480, loss: 0.0002981330908369273
step: 490, loss: 0.016125451773405075
step: 500, loss: 3.8623984437435865e-05
step: 510, loss: 0.15395165979862213
step: 520, loss: 8.98751022759825e-05
step: 530, loss: 2.4251161448773928e-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=0.9469234382339127, f1=0.9398601398601399,
best_f1=0.9391143911439114
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 [00:16, 347.40it/s]
load_f1 = 0.9456066945606694

```



```
real_f1 = 0.9441860465116279
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 [00:12, 363.94it/s]
```