HLow_10_1_2_distilbert

August 10, 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/Snippext_public
     %cd Snippext_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 | 19.19 MiB/s, done.
Resolving deltas: 100% (142/142), done.
Checking out files: 100% (128/128), done.
/content/ditto
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting gensim==3.8.1
  Downloading gensim-3.8.1-cp37-cp37m-manylinux1_x86_64.whl (24.2 MB)
     |                        | 24.2 MB 1.5 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
     |                        | 14.5 MB 52.1 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
     |                        | 689 kB 70.3 MB/s
Collecting scipy==1.3.2
  Downloading scipy-1.3.2-cp37-cp37m-manylinux1_x86_64.whl (25.2 MB)
     |                        | 25.2 MB 1.2 MB/s
Collecting sentencepiece==0.1.85
  Downloading sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
     |                        | 1.0 MB 8.3 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 57.8 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 | 20.83 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 4.9 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 76.2 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
     |                      | 753.4 MB 7.0 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
     |                      | 56 kB 6.4 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
     |                      | 884 kB 54.4 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 61.3 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)
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 44.9 MB/s
Collecting plac<1.2.0,>=0.9.6
  Downloading plac-1.1.3-py2.py3-none-any.whl (20 kB)
Requirement already satisfied: 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: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (2.0.6)
Requirement already satisfied: 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: 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 63.6 MB/s
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 catalogue<1.1.0,>=0.0.7
  Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
Collecting srsly<1.1.0,>=0.1.0
  Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
     |                    | 184 kB 68.0 MB/s
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.7/dist-packages (from sklearn==0.0->-r requirements.txt
(line 6)) (1.0.2)
Requirement already satisfied: protobuf>=3.8.0 in /usr/local/lib/python3.7/dist-
packages (from tensorboardX==2.0->-r requirements.txt (line 8)) (3.17.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-
packages (from transformers==3.1.0->-r requirements.txt (line 11)) (3.7.1)
```

Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from transformers==3.1.0->-r requirements.txt (line 11)) (21.3)
Collecting sacremoses
  Downloading sacremoses-0.0.53.tar.gz (880 kB)
     | | 880 kB 57.4 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 40.1 MB/s
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: 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: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (1.24.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line
4)) (2.10)
Requirement already satisfied: 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=1449927
sha256=ac757bb7f182a60e8e9684057bf4f87b4da7a7a865df889677b6d84ea4b264b1
  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=e47c1e8809f3b96a45868e830c4549954e4a0f2036ad6b31d6196e935460683f
  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: 10117, done.
remote: Counting objects: 100% (233/233), done.
remote: Compressing objects: 100% (165/165), done.
remote: Total 10117 (delta 119), reused 152 (delta 62), pack-reused 9884
Receiving objects: 100% (10117/10117), 14.97 MiB | 23.65 MiB/s, done.
Resolving deltas: 100% (6927/6927), 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 writeable
Created temporary directory: /tmp/pip-ephem-wheel-cache-a18aj7ii
Created temporary directory: /tmp/pip-req-tracker-6qo_v3vb
Initialized build tracking at /tmp/pip-req-tracker-6qo_v3vb
Created build tracker: /tmp/pip-req-tracker-6qo_v3vb
Entered build tracker: /tmp/pip-req-tracker-6qo_v3vb
```

```
Created temporary directory: /tmp/pip-install-ohgxpotj
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-u3li303z
  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-6qo_v3vb'
    Running setup.py (path:/tmp/pip-req-build-u3li303z/setup.py) egg_info for
package from file:///content/ditto/apex
    Created temporary directory: /tmp/pip-pip-egg-info-pj8r3rj6
    Running command python setup.py egg_info


    torch.__version__  = 1.4.0


    running egg_info
    creating /tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-info
    writing /tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-info/PKG-INFO
    writing dependency_links to /tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-
info/dependency_links.txt
    writing top-level names to /tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-
info/top_level.txt
    writing manifest file '/tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-
info/SOURCES.txt'
    adding license file 'LICENSE'
    writing manifest file '/tmp/pip-pip-egg-info-pj8r3rj6/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-u3li303z 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-6qo_v3vb'
Created temporary directory: /tmp/pip-unpack-yr01j4h0
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-zcs_7lw9
  Building wheel for apex (setup.py) …   Destination directory:
/tmp/pip-wheel-zcs_7lw9
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = '"'"'/tmp/pip-req-build-u3li303z/setup.py'"'"';
```

```
__file__='"'"'/tmp/pip-req-build-u3li303z/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-zcs_7lw9


  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/amp
  copying apex/amp/scaler.py -> build/lib/apex/amp
  copying apex/amp/_initialize.py -> build/lib/apex/amp
  copying apex/amp/handle.py -> build/lib/apex/amp
  copying apex/amp/opt.py -> build/lib/apex/amp
  copying apex/amp/_amp_state.py -> build/lib/apex/amp
  copying apex/amp/__init__.py -> build/lib/apex/amp
  copying apex/amp/wrap.py -> build/lib/apex/amp
  copying apex/amp/rnn_compat.py -> build/lib/apex/amp
  copying apex/amp/_process_optimizer.py -> build/lib/apex/amp
  copying apex/amp/utils.py -> build/lib/apex/amp
  copying apex/amp/amp.py -> build/lib/apex/amp
  copying apex/amp/frontend.py -> build/lib/apex/amp
  copying apex/amp/compat.py -> build/lib/apex/amp
  copying apex/amp/__version__.py -> build/lib/apex/amp
  creating build/lib/apex/optimizers
  copying apex/optimizers/fused_lamb.py -> build/lib/apex/optimizers
  copying apex/optimizers/fused_adagrad.py -> build/lib/apex/optimizers
  copying apex/optimizers/__init__.py -> build/lib/apex/optimizers
  copying apex/optimizers/fused_mixed_precision_lamb.py ->
build/lib/apex/optimizers
  copying apex/optimizers/fused_novograd.py -> build/lib/apex/optimizers
  copying apex/optimizers/fused_sgd.py -> build/lib/apex/optimizers
  copying apex/optimizers/fused_adam.py -> build/lib/apex/optimizers
  creating build/lib/apex/contrib
  copying apex/contrib/__init__.py -> build/lib/apex/contrib
  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/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/mlp.py -> build/lib/apex/mlp
copying apex/mlp/__init__.py -> build/lib/apex/mlp
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/parallel
copying apex/parallel/LARC.py -> build/lib/apex/parallel
copying apex/parallel/multiproc.py -> build/lib/apex/parallel
copying apex/parallel/__init__.py -> build/lib/apex/parallel
copying apex/parallel/distributed.py -> build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm.py -> build/lib/apex/parallel
copying apex/parallel/sync_batchnorm_kernel.py -> build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/lib/apex/parallel
copying apex/parallel/sync_batchnorm.py -> build/lib/apex/parallel
creating build/lib/apex/transformer
copying apex/transformer/parallel_state.py -> build/lib/apex/transformer
copying apex/transformer/enums.py -> build/lib/apex/transformer
copying apex/transformer/__init__.py -> build/lib/apex/transformer
copying apex/transformer/microbatches.py -> build/lib/apex/transformer
copying apex/transformer/utils.py -> build/lib/apex/transformer
copying apex/transformer/log_util.py -> build/lib/apex/transformer
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/loss_scaler.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/fp16util.py -> build/lib/apex/fp16_utils
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/amp/lists
copying apex/amp/lists/functional_overrides.py -> build/lib/apex/amp/lists
copying apex/amp/lists/tensor_overrides.py -> build/lib/apex/amp/lists
copying apex/amp/lists/__init__.py -> build/lib/apex/amp/lists
copying apex/amp/lists/torch_overrides.py -> build/lib/apex/amp/lists
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/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_self_multihead_attn_norm_add_func.py
-> build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/self_multihead_attn_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/__init__.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/encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
  copying apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
  creating build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_lamb.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fp16_optimizer.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/__init__.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/distributed_fused_lamb.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/distributed_fused_adam.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_sgd.py ->
build/lib/apex/contrib/optimizers
  copying apex/contrib/optimizers/fused_adam.py ->
build/lib/apex/contrib/optimizers
  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/groupbn
  copying apex/contrib/groupbn/batch_norm.py -> build/lib/apex/contrib/groupbn
  copying apex/contrib/groupbn/__init__.py -> build/lib/apex/contrib/groupbn
  creating build/lib/apex/contrib/transducer
  copying apex/contrib/transducer/transducer.py ->
build/lib/apex/contrib/transducer
  copying apex/contrib/transducer/__init__.py ->
build/lib/apex/contrib/transducer
  creating 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_halo_exchange_module_tests.py ->
build/lib/apex/contrib/peer_memory
  copying apex/contrib/peer_memory/peer_memory.py ->
build/lib/apex/contrib/peer_memory
  copying apex/contrib/peer_memory/__init__.py ->
build/lib/apex/contrib/peer_memory
  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/sparse_masklib.py ->
build/lib/apex/contrib/sparsity
  copying apex/contrib/sparsity/asp.py -> build/lib/apex/contrib/sparsity
  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/bottleneck
  copying apex/contrib/bottleneck/bottleneck.py ->
build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/bottleneck_module_test.py ->
build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/__init__.py ->
build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/test.py -> build/lib/apex/contrib/bottleneck
  copying apex/contrib/bottleneck/halo_exchangers.py ->
build/lib/apex/contrib/bottleneck
  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/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/layer_norm
  copying apex/contrib/layer_norm/__init__.py ->
build/lib/apex/contrib/layer_norm
  copying apex/contrib/layer_norm/layer_norm.py ->
build/lib/apex/contrib/layer_norm
  creating build/lib/apex/contrib/index_mul_2d
  copying apex/contrib/index_mul_2d/index_mul_2d.py ->
build/lib/apex/contrib/index_mul_2d
  copying apex/contrib/index_mul_2d/__init__.py ->
build/lib/apex/contrib/index_mul_2d
```

```
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
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/layers
copying apex/transformer/layers/__init__.py ->
build/lib/apex/transformer/layers
copying apex/transformer/layers/layer_norm.py ->
build/lib/apex/transformer/layers
creating build/lib/apex/transformer/functional
copying apex/transformer/functional/fused_softmax.py ->
build/lib/apex/transformer/functional
copying apex/transformer/functional/__init__.py ->
build/lib/apex/transformer/functional
creating 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/__init__.py ->
build/lib/apex/transformer/pipeline_parallel
copying apex/transformer/pipeline_parallel/_timers.py ->
build/lib/apex/transformer/pipeline_parallel
copying apex/transformer/pipeline_parallel/utils.py ->
build/lib/apex/transformer/pipeline_parallel
creating 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/__init__.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/memory.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/random.py ->
build/lib/apex/transformer/tensor_parallel
copying apex/transformer/tensor_parallel/utils.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/mappings.py ->
```

```
build/lib/apex/transformer/tensor_parallel
  creating build/lib/apex/transformer/testing
  copying apex/transformer/testing/distributed_test_base.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/global_vars.py ->
build/lib/apex/transformer/testing
  copying apex/transformer/testing/commons.py ->
build/lib/apex/transformer/testing
  copying apex/transformer/testing/standalone_gpt.py ->
build/lib/apex/transformer/testing
  copying apex/transformer/testing/arguments.py ->
build/lib/apex/transformer/testing
  copying apex/transformer/testing/standalone_transformer_lm.py ->
build/lib/apex/transformer/testing
  creating build/lib/apex/transformer/_data
  copying apex/transformer/_data/__init__.py -> build/lib/apex/transformer/_data
  copying apex/transformer/_data/_batchsampler.py ->
build/lib/apex/transformer/_data
  creating 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/common.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/fwd_bwd_no_pipelining.py
-> build/lib/apex/transformer/pipeline_parallel/schedules
  copying apex/transformer/pipeline_parallel/schedules/__init__.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
  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/amp
  copying build/lib/apex/amp/scaler.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_initialize.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/handle.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/opt.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/__init__.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/rnn_compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_process_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  creating build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/functional_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/tensor_overrides.py ->
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/torch_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/utils.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/amp.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/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
  creating build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_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/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_mixed_precision_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_novograd.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_sgd.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
  creating build/bdist.linux-x86_64/wheel/apex/contrib
  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/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_self_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_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/__init__.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/encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
  copying build/lib/apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
  copying build/lib/apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
  copying
build/lib/apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
  creating build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  copying build/lib/apex/contrib/optimizers/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  copying build/lib/apex/contrib/optimizers/fp16_optimizer.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/distributed_fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  copying build/lib/apex/contrib/optimizers/distributed_fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  copying build/lib/apex/contrib/optimizers/fused_sgd.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  copying build/lib/apex/contrib/optimizers/fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
  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/groupbn
  copying build/lib/apex/contrib/groupbn/batch_norm.py ->
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/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib
```

```
creating 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
copying build/lib/apex/contrib/transducer/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/transducer
creating 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_halo_exchange_module_tests.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
copying build/lib/apex/contrib/peer_memory/peer_memory.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
copying build/lib/apex/contrib/peer_memory/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
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
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/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
copying build/lib/apex/contrib/sparsity/sparse_masklib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
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/conv_bias_relu
copying build/lib/apex/contrib/conv_bias_relu/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
copying build/lib/apex/contrib/conv_bias_relu/conv_bias_relu.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
creating build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
copying build/lib/apex/contrib/bottleneck/bottleneck.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/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
  copying build/lib/apex/contrib/bottleneck/test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
  copying build/lib/apex/contrib/bottleneck/halo_exchangers.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
  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/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
  creating build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
  copying build/lib/apex/contrib/layer_norm/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
  copying build/lib/apex/contrib/layer_norm/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
  creating build/bdist.linux-x86_64/wheel/apex/contrib/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
  copying build/lib/apex/contrib/index_mul_2d/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
  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/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
  copying build/lib/apex/__init__.py -> build/bdist.linux-x86_64/wheel/apex
  creating build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/mlp.py -> build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/mlp
  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
  copying build/lib/apex/_autocast_utils.py ->
build/bdist.linux-x86_64/wheel/apex
  creating 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/multiproc.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/distributed.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
  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_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  creating build/bdist.linux-x86_64/wheel/apex/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
  creating build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/parallel_state.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/fused_softmax.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/enums.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/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/p2p_communication.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
  creating
```

```
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelin
ing_with_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/schedules/fwd_bwd_pipelin
ing_without_interleaving.py ->
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/__init__.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
  copying build/lib/apex/transformer/pipeline_parallel/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/microbatches.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/cross_entropy.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/data.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/memory.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/random.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/layers.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
  creating 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/__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/global_vars.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/standalone_gpt.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/arguments.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_transformer_lm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/utils.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/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/_batchsampler.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/_data
  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/loss_scaler.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16util.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  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
  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-zcs_7lw9/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/_batchsampler.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=298244
sha256=723b7fb32ac3f209b8397b10a5c716178c5e1c7cc3d27cf8604b9a8a9752eaf9
  Stored in directory: /tmp/pip-ephem-wheel-cache-a18aj7ii/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-6qo_v3vb'
/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 5.0 MB/s
Collecting awscli
  Downloading awscli-1.25.48-py3-none-any.whl (3.9 MB)
     | | 3.9 MB 37.1 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
     | | 79 kB 10.7 MB/s
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Requirement already satisfied: PyYAML<5.5,>=3.10 in
/usr/local/lib/python3.7/dist-packages (from awscli) (3.13)
Collecting botocore==1.27.48
  Downloading botocore-1.27.48-py3-none-any.whl (9.0 MB)
     | | 9.0 MB 47.4 MB/s
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting docutils<0.17,>=0.10
  Downloading docutils-0.16-py2.py3-none-any.whl (548 kB)
     | | 548 kB 7.9 MB/s
Collecting jmespath<2.0.0,>=0.7.1

```
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in
/usr/local/lib/python3.7/dist-packages (from botocore==1.27.48->awscli) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil<3.0.0,>=2.1->botocore==1.27.48->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-1.25.48 botocore-1.27.48 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: numpy in /usr/local/lib/python3.7/dist-
packages (from torch==1.7.1) (1.19.2)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
Installing collected packages: torch
  Attempting uninstall: torch
    Found existing installation: torch 1.4.0
```

```
     Uninstalling torch-1.4.0:
        Successfully uninstalled torch-1.4.0
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.
torchvision 0.13.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.
torchtext 0.13.0 requires torch==1.12.0, but you have torch 1.7.1 which is
incompatible.
torchaudio 0.12.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.
Successfully installed torch-1.7.1
```

0.3   importa config.json

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

```
Cloning into 'IC'…
remote: Enumerating objects: 1014, done.
remote: Counting objects: 100% (84/84), done.
remote: Compressing objects: 100% (74/74), done.
remote: Total 1014 (delta 27), reused 7 (delta 0), pack-reused 930
Receiving objects: 100% (1014/1014), 254.12 MiB | 30.90 MiB/s, done.
Resolving deltas: 100% (611/611), done.
Checking out files: 100% (1284/1284), 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/HLow_10_1_2/configs.json /content/ditto/
```

1 DA

1.1 DA STRUCTURED

1.1.1 Beer - Running the matcher

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

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

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

```
Downloading: 100% 442/442 [00:00<00:00, 598kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 24.0MB/s]
Downloading: 100% 268M/268M [00:05<00:00, 45.7MB/s]
step: 0, loss: 0.8665435910224915
epoch 1: dev_f1=0.27184466019417475, f1=0.2666666666666667,
best_f1=0.2666666666666667
step: 0, loss: 0.3802330791950226
epoch 2: dev_f1=0.2692307692307693, f1=0.2692307692307693,
best_f1=0.2666666666666667
step: 0, loss: 0.3539169132709503
epoch 3: dev_f1=0.358974358974359, f1=0.3466666666666666,
best_f1=0.3466666666666666
```

```
step: 0, loss: 0.37102386355400085
epoch 4: dev_f1=0.3888888888888889, f1=0.35294117647058826,
best_f1=0.35294117647058826
step: 0, loss: 0.2871374785900116
epoch 5: dev_f1=0.4666666666666667, f1=0.3728813559322034,
best_f1=0.3728813559322034
step: 0, loss: 0.2949204742908478
epoch 6: dev_f1=0.5, f1=0.43999999999999995, best_f1=0.43999999999999995
step: 0, loss: 0.3194042444229126
epoch 7: dev_f1=0.49122807017543857, f1=0.42857142857142855,
best_f1=0.43999999999999995
step: 0, loss: 0.44439104199409485
epoch 8: dev_f1=0.5, f1=0.4571428571428571, best_f1=0.43999999999999995
step: 0, loss: 0.20966701209545135
epoch 9: dev_f1=0.47058823529411764, f1=0.4864864864864865,
best_f1=0.43999999999999995
step: 0, loss: 0.2509379982948303
epoch 10: dev_f1=0.5333333333333333, f1=0.375, best_f1=0.375
step: 0, loss: 0.1673606038093567
epoch 11: dev_f1=0.5925925925925927, f1=0.43750000000000006,
best_f1=0.43750000000000006
step: 0, loss: 0.20909132063388824
epoch 12: dev_f1=0.5925925925925927, f1=0.4137931034482759,
best_f1=0.43750000000000006
step: 0, loss: 0.11244281381368637
epoch 13: dev_f1=0.56, f1=0.4444444444444445, best_f1=0.43750000000000006
step: 0, loss: 0.20149601995944977
epoch 14: dev_f1=0.5405405405405405, f1=0.5000000000000001,
best_f1=0.43750000000000006
step: 0, loss: 0.16780392825603485
epoch 15: dev_f1=0.5405405405405405, f1=0.5000000000000001,
best_f1=0.43750000000000006
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

```python
# 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 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: 0.8140081167221069
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4691087007522583
step: 20, loss: 0.573684573173523
step: 30, loss: 0.3601445257663727
step: 40, loss: 0.13057567179203033
step: 50, loss: 0.09734035283327103
step: 60, loss: 0.07255251705646515
step: 70, loss: 0.03976398706436157
step: 80, loss: 0.09027788788080215
step: 90, loss: 0.020716411992907524
step: 100, loss: 0.14662455022335052
step: 110, loss: 0.012072542682290077
step: 120, loss: 0.006119802128523588
step: 130, loss: 0.004927795846015215
step: 140, loss: 0.011790983378887177
step: 150, loss: 0.13221679627895355
step: 160, loss: 0.05562286078929901
step: 170, loss: 0.030703557655215263
step: 180, loss: 0.004017230123281479
step: 190, loss: 0.015612353570759296
step: 200, loss: 0.00747125456109643
step: 210, loss: 0.008367745205760002
step: 220, loss: 0.0016295070527121425
step: 230, loss: 0.0027302971575409174
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9887387387387387, f1=0.9853438556933484,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003924786113202572
step: 10, loss: 0.006745054852217436
step: 20, loss: 0.0023281702306121588
step: 30, loss: 0.0017087801825255156
step: 40, loss: 0.0036225044168531895
step: 50, loss: 0.004256185609847307
step: 60, loss: 0.003418585518375039
step: 70, loss: 0.046795327216386795
step: 80, loss: 0.003734086873009801
step: 90, loss: 0.03148221597075462
step: 100, loss: 0.020937712863087654
step: 110, loss: 0.13409662246704102
step: 120, loss: 0.0019515762105584145
step: 130, loss: 0.0025126663967967033
step: 140, loss: 0.03557892516255379
step: 150, loss: 0.0056621478870511055
step: 160, loss: 0.005959208123385906
step: 170, loss: 0.005203690379858017
step: 180, loss: 0.0017933952622115612
step: 190, loss: 0.14926792681217194
step: 200, loss: 0.006226774305105209
step: 210, loss: 0.08745919913053513
step: 220, loss: 0.0009505777852609754
step: 230, loss: 0.042296912521123886
/usr/local/lib/python3.7/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=0.9819004524886877, f1=0.9720044792833147,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03174508363008499
step: 10, loss: 0.02714351750910282
step: 20, loss: 0.015727778896689415
step: 30, loss: 0.0027549422811716795
step: 40, loss: 0.0051818666979670525
step: 50, loss: 0.0022847475484013557
step: 60, loss: 0.0006736161303706467
step: 70, loss: 0.0006128028035163879
step: 80, loss: 0.0029125716537237167
step: 90, loss: 0.0007707241456955671
```

```
step: 100, loss: 0.00047660747077316046
step: 110, loss: 0.001599695417098701
step: 120, loss: 0.0005406284471973777
step: 130, loss: 0.00034847293864004314
step: 140, loss: 0.08608581125736237
step: 150, loss: 0.0012354105710983276
step: 160, loss: 0.0012150179827585816
step: 170, loss: 0.06447073072195053
step: 180, loss: 0.0022484706714749336
step: 190, loss: 0.0022759470157325268
step: 200, loss: 0.0008143965387716889
step: 210, loss: 0.012598316185176373
step: 220, loss: 0.0008579265559092164
step: 230, loss: 0.07844040542840958
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9932432432432432, f1=0.9865168539325843,
best_f1=0.9865168539325843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003496070858091116
step: 10, loss: 0.0008098743273876607
step: 20, loss: 0.009481679648160934
step: 30, loss: 0.0008121097926050425
step: 40, loss: 0.0015247537521645427
step: 50, loss: 0.0013154476182533529
step: 60, loss: 0.00042037206003442407
step: 70, loss: 0.031531352549791336
step: 80, loss: 0.030479835346341133
step: 90, loss: 0.026737557724118233
step: 100, loss: 0.0012840427225455642
step: 110, loss: 0.0058157797902822495
step: 120, loss: 0.004627666901797056
step: 130, loss: 0.0011956279631704092
step: 140, loss: 0.004154601134359837
step: 150, loss: 0.0011918412055820227
step: 160, loss: 0.000450229796115309
step: 170, loss: 0.004938116297125816
step: 180, loss: 0.030866356566548347
step: 190, loss: 0.009613231755793095
step: 200, loss: 0.00052018987480551
step: 210, loss: 0.00021609259420074522
```

```
step: 220, loss: 0.00025561946677044034
step: 230, loss: 0.0002388609282206744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9943757030371203, f1=0.9898762654668166,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003965660580433905
step: 10, loss: 0.00021281153021845967
step: 20, loss: 0.00018194818403571844
step: 30, loss: 0.062122333379483223
step: 40, loss: 0.0005223894258961082
step: 50, loss: 0.028208250179886818
step: 60, loss: 0.008894076570868492
step: 70, loss: 0.0004532143648248166
step: 80, loss: 0.0006877603591419756
step: 90, loss: 0.0008110228809528053
step: 100, loss: 0.0020975107327103615
step: 110, loss: 0.00049778574611999164
step: 120, loss: 0.08676525950431824
step: 130, loss: 0.05072388797998428
step: 140, loss: 0.0003157676837872714
step: 150, loss: 0.0004381255421321839
step: 160, loss: 0.006636145990341902
step: 170, loss: 0.052181608974933624
step: 180, loss: 0.0006521171890199184
step: 190, loss: 0.0008722394122742116
step: 200, loss: 0.029845040291547775
step: 210, loss: 0.0026362519711256027
step: 220, loss: 0.026545600965619087
step: 230, loss: 0.0007707682088948786
/usr/local/lib/python3.7/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=0.9864864864864865, f1=0.9865470852017937,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010346604976803064
step: 10, loss: 0.0010197340743616223
```

```
step: 20, loss: 0.0018333137268200517
step: 30, loss: 0.00024323402612935752
step: 40, loss: 0.0005471982876770198
step: 50, loss: 0.19596321880817413
step: 60, loss: 0.017937662079930305
step: 70, loss: 0.0008400179794989526
step: 80, loss: 0.0005325258243829012
step: 90, loss: 0.00029691687086597085
step: 100, loss: 0.00090765644563362
step: 110, loss: 0.035590700805187225
step: 120, loss: 0.000637534074485302
step: 130, loss: 0.0007954038446769118
step: 140, loss: 0.0012190481647849083
step: 150, loss: 0.0002245923096779734
step: 160, loss: 0.017515532672405243
step: 170, loss: 0.000697230629157275
step: 180, loss: 0.0004940746584907174
step: 190, loss: 0.0005247552180662751
step: 200, loss: 0.0007453255821019411
step: 210, loss: 0.00028412052779458463
step: 220, loss: 0.0003098613815382123
step: 230, loss: 0.00027718936325982213
/usr/local/lib/python3.7/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=0.9898305084745763, f1=0.9853107344632768,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06200769171118736
step: 10, loss: 0.00011114829647121951
step: 20, loss: 0.00038673859671168957
step: 30, loss: 0.0005642006290145218
step: 40, loss: 0.0017680919263511896
step: 50, loss: 0.0001600373216206208
step: 60, loss: 0.0008658870356157422
step: 70, loss: 0.0013927583349868655
step: 80, loss: 0.0022567531559616327
step: 90, loss: 0.0003163551155012101
step: 100, loss: 0.00107599759940058
step: 110, loss: 0.002026587724685669
step: 120, loss: 0.0008117405814118683
step: 130, loss: 0.00045165876508690417
step: 140, loss: 0.00015254103345796466
step: 150, loss: 0.0006609727279283106
step: 160, loss: 0.00015017374244052917
step: 170, loss: 9.719579247757792e-05
```

```
step: 180, loss: 0.00039266873500309885
step: 190, loss: 0.0002884742571040988
step: 200, loss: 0.0004898349870927632
step: 210, loss: 0.00038836695603094995
step: 220, loss: 0.0009762306581251323
step: 230, loss: 0.02067333459854126
/usr/local/lib/python3.7/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=0.9932126696832579, f1=0.9876265466816648,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024096090346574783
step: 10, loss: 0.03748241439461708
step: 20, loss: 0.00038177979877218604
step: 30, loss: 0.0001308679929934442
step: 40, loss: 7.155393541324884e-05
step: 50, loss: 0.0005258244927972555
step: 60, loss: 9.150036203209311e-05
step: 70, loss: 0.00010643043060554191
step: 80, loss: 0.0003740888787433505
step: 90, loss: 0.04577833414077759
step: 100, loss: 0.00037005243939347565
step: 110, loss: 0.0027894426602870226
step: 120, loss: 0.0005223756888881326
step: 130, loss: 0.013579479418694973
step: 140, loss: 9.353381028631702e-05
step: 150, loss: 0.00013070114073343575
step: 160, loss: 0.00019391310343053192
step: 170, loss: 0.00012063724716426805
step: 180, loss: 0.00229084026068449
step: 190, loss: 0.0006157042225822806
step: 200, loss: 0.011358052492141724
step: 210, loss: 0.00015097259893082082
step: 220, loss: 0.0007933764136396348
step: 230, loss: 0.009090942330658436
/usr/local/lib/python3.7/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=0.9888641425389755, f1=0.9790055248618784,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004030887794215232
step: 10, loss: 0.00035966746509075165
```

```
step: 20, loss: 0.0015762749826535583
step: 30, loss: 0.0018771030008792877
step: 40, loss: 0.0007729872013442218
step: 50, loss: 0.0001679630804574117
step: 60, loss: 0.0001479578932048753
step: 70, loss: 0.00027197948656976223
step: 80, loss: 0.03315107524394989
step: 90, loss: 0.000757219095248729
step: 100, loss: 0.0004576731298584491
step: 110, loss: 0.0002554154198151082
step: 120, loss: 0.02303021028637886
step: 130, loss: 0.00011082296259701252
step: 140, loss: 0.0003475246485322714
step: 150, loss: 8.336368773598224e-05
step: 160, loss: 0.000808577227871865
step: 170, loss: 0.00010392387048341334
step: 180, loss: 7.396802539005876e-05
step: 190, loss: 4.5484088332159445e-05
step: 200, loss: 8.540426642866805e-05
step: 210, loss: 6.327262963168323e-05
step: 220, loss: 0.055633027106523514
step: 230, loss: 0.008754286915063858
/usr/local/lib/python3.7/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=0.988814317673378, f1=0.9844097995545658,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.891186735127121e-05
step: 10, loss: 0.00029154546479168761
step: 20, loss: 0.0002240599278593436
step: 30, loss: 7.838533201720566e-05
step: 40, loss: 0.00013121427036821842
step: 50, loss: 0.0004683865117840469
step: 60, loss: 0.071903094664931488
step: 70, loss: 9.24423075048253e-05
step: 80, loss: 0.00023663384490646422
step: 90, loss: 6.382525316439569e-05
step: 100, loss: 0.00010227365419268608
step: 110, loss: 0.0017199160065501928
step: 120, loss: 0.016702765598893166
step: 130, loss: 9.333579509984702e-05
step: 140, loss: 0.020493244752287865
step: 150, loss: 7.029173139017075e-05
step: 160, loss: 0.0005408247234299779
step: 170, loss: 7.951271982165053e-05
```

```
step: 180, loss: 0.0001234432857017964
step: 190, loss: 0.018397558480501175
step: 200, loss: 7.04864869476296e-05
step: 210, loss: 0.00010615297651384026
step: 220, loss: 0.000896969111636281
step: 230, loss: 7.964618998812512e-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=0.9909297052154196, f1=0.9898305084745763,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.805252633057535e-05
step: 10, loss: 0.00018362478294875473
step: 20, loss: 4.092346352990717e-05
step: 30, loss: 7.355007983278483e-05
step: 40, loss: 0.0012074876576662064
step: 50, loss: 0.0002560083521530032
step: 60, loss: 0.00014268254744820297
step: 70, loss: 0.00021165366342756897
step: 80, loss: 0.0001365126227028668
step: 90, loss: 6.950860552024096e-05
step: 100, loss: 0.000146016594953835
step: 110, loss: 0.00012332179176155478
step: 120, loss: 0.00019958085613325238
step: 130, loss: 0.0006382521241903305
step: 140, loss: 6.749866588506848e-05
step: 150, loss: 4.760877345688641e-05
step: 160, loss: 7.95002852100879e-05
step: 170, loss: 0.019342467188835144
step: 180, loss: 9.600319754099473e-05
step: 190, loss: 0.000404528109356761
step: 200, loss: 6.44908650428988e-05
step: 210, loss: 4.480525240069255e-05
step: 220, loss: 5.904526187805459e-05
step: 230, loss: 0.014148738235235214
/usr/local/lib/python3.7/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=0.9909297052154196, f1=0.9898305084745763,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0025627745781093836
step: 10, loss: 6.93256442900747e-05
```

```
step: 20, loss: 0.03594515472650528
step: 30, loss: 8.303968206746504e-05
step: 40, loss: 8.018094376893714e-05
step: 50, loss: 6.640706851612777e-05
step: 60, loss: 0.008348433300852776
step: 70, loss: 6.218514317879453e-05
step: 80, loss: 8.759225602261722e-05
step: 90, loss: 7.169948366936296e-05
step: 100, loss: 5.124327799421735e-05
step: 110, loss: 6.392419891199097e-05
step: 120, loss: 6.303418194875121e-05
step: 130, loss: 0.00014706410001963377
step: 140, loss: 3.896120688295923e-05
step: 150, loss: 6.680107617285103e-05
step: 160, loss: 0.06426241993904114
step: 170, loss: 5.2846738981315866e-05
step: 180, loss: 5.9907739341724664e-05
step: 190, loss: 0.001432760269381106
step: 200, loss: 0.0046310862526297577
step: 210, loss: 4.6808992919977754e-05
step: 220, loss: 0.050650645047426224
step: 230, loss: 7.482584624085575e-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=0.9854748603351955,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.158021303941496e-05
step: 10, loss: 0.0001009945772239007
step: 20, loss: 0.000170640429132618
step: 30, loss: 0.020292211323976517
step: 40, loss: 0.0003317256923764944
step: 50, loss: 5.4929365433054045e-05
step: 60, loss: 5.9318099374650046e-05
step: 70, loss: 7.35939247533679e-05
step: 80, loss: 0.0007099208887666464
step: 90, loss: 3.6808127333642915e-05
step: 100, loss: 5.423348193289712e-05
step: 110, loss: 4.5006931031821296e-05
step: 120, loss: 0.00015318595978897065
step: 130, loss: 0.0001064584357663989
step: 140, loss: 0.0001063078161678277
step: 150, loss: 0.011652717366814613
step: 160, loss: 4.574408012558706e-05
step: 170, loss: 7.937143527669832e-05
```

```
step: 180, loss: 5.809607318951748e-05
step: 190, loss: 3.407019903534092e-05
step: 200, loss: 4.4335498387226835e-05
step: 210, loss: 5.8933899708790705e-05
step: 220, loss: 3.399166598683223e-05
step: 230, loss: 3.254681723774411e-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=0.9920724801812004, f1=0.9898305084745763,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.828765228623524e-05
step: 10, loss: 2.427318941045087e-05
step: 20, loss: 0.01605171710252762
step: 30, loss: 8.952963253250346e-05
step: 40, loss: 2.6932862965622917e-05
step: 50, loss: 0.00012139269529143348
step: 60, loss: 3.3455136872362345e-05
step: 70, loss: 5.534591764444485e-05
step: 80, loss: 5.866812716703862e-05
step: 90, loss: 7.07268191035837e-05
step: 100, loss: 0.017439967021346092
step: 110, loss: 4.1185365262208506e-05
step: 120, loss: 0.0001746136840665713
step: 130, loss: 8.36335530038923e-05
step: 140, loss: 5.490730836754665e-05
step: 150, loss: 6.80223383824341e-05
step: 160, loss: 2.6050154701806605e-05
step: 170, loss: 3.808440669672564e-05
step: 180, loss: 5.502522617462091e-05
step: 190, loss: 0.011664364486932755
step: 200, loss: 3.036764064745512e-05
step: 210, loss: 0.012486412189900875
step: 220, loss: 5.027119550504722e-05
step: 230, loss: 2.4698076231288724e-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=0.9920724801812004, f1=0.9898305084745763,
best_f1=0.9898762654668166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.688462649937719e-05
step: 10, loss: 6.489168299594894e-05
```

```
step: 20, loss: 5.3889103583060205e-05
step: 30, loss: 0.00018402651767246425
step: 40, loss: 6.566184310941026e-05
step: 50, loss: 8.359304047189653e-05
step: 60, loss: 3.01103664241964e-05
step: 70, loss: 6.387372559402138e-05
step: 80, loss: 0.0260853860527277
step: 90, loss: 2.225443131464999e-05
step: 100, loss: 4.634038850781508e-05
step: 110, loss: 4.717438423540443e-05
step: 120, loss: 3.984742579632439e-05
step: 130, loss: 4.7635476221330464e-05
step: 140, loss: 6.226267578313127e-05
step: 150, loss: 9.921252058120444e-05
step: 160, loss: 0.00010403990017948672
step: 170, loss: 2.805786789394915e-05
step: 180, loss: 0.00010874723375309259
step: 190, loss: 0.011390815488994122
step: 200, loss: 5.261966725811362e-05
step: 210, loss: 0.023808341473340988
step: 220, loss: 4.588154479279183e-05
step: 230, loss: 8.519646507920697e-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=0.9920724801812004, f1=0.9887133182844244,
best_f1=0.9898762654668166
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
```

```
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:06, 358.10it/s]
load_f1 = 0.9932584269662922
real_f1 = 0.9932584269662922
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 432.35it/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, 628kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 21.8MB/s]
Downloading: 100% 268M/268M [00:05<00:00, 46.8MB/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: 0.7865638136863708
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4682918190956116
step: 20, loss: 0.4827061593532562

```
step: 30, loss: 0.3883017599582672
step: 40, loss: 0.2758939564228058
step: 50, loss: 0.18208491802215576
step: 60, loss: 0.08209817856550217
step: 70, loss: 0.07624722272157669
step: 80, loss: 0.14013971388339996
step: 90, loss: 0.10397667437791824
step: 100, loss: 0.3243309557437897
step: 110, loss: 0.04671727493405342
step: 120, loss: 0.03644358739256859
step: 130, loss: 0.03123968467116356
step: 140, loss: 0.18547484278678894
step: 150, loss: 0.053430091589689255
step: 160, loss: 0.12719473242759705
step: 170, loss: 0.06802865862846375
step: 180, loss: 0.09555365145206451
step: 190, loss: 0.029369976371526718
step: 200, loss: 0.1218835860490799
step: 210, loss: 0.07901956886053085
step: 220, loss: 0.04736173152923584
step: 230, loss: 0.06407322734594345
step: 240, loss: 0.05244471877813339
step: 250, loss: 0.025111282244324684
step: 260, loss: 0.07302247732877731
step: 270, loss: 0.018297113478183746
step: 280, loss: 0.05943707376718521
step: 290, loss: 0.10288722813129425
step: 300, loss: 0.13079099357128143
step: 310, loss: 0.05667676776647568
step: 320, loss: 0.06492401659488678
step: 330, loss: 0.0864153727889061
step: 340, loss: 0.13785114884376526
step: 350, loss: 0.11077739298343658
step: 360, loss: 0.06958940625190735
step: 370, loss: 0.12143511325120926
step: 380, loss: 0.1637454330921173
step: 390, loss: 0.02510562166571617
step: 400, loss: 0.008467624895274639
step: 410, loss: 0.01065621618181467
step: 420, loss: 0.010898482985794544
step: 430, loss: 0.0879179984331131
step: 440, loss: 0.06496266275644302
step: 450, loss: 0.013497811742126942
step: 460, loss: 0.08904408663511276
step: 470, loss: 0.2696780264377594
step: 480, loss: 0.2675829231739044
step: 490, loss: 0.03649287670850754
step: 500, loss: 0.010642669163644314
```

```
step: 510, loss: 0.03548290580511093
step: 520, loss: 0.007174062542617321
step: 530, loss: 0.07462003082036972
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9348127600554785, f1=0.9351851851851852,
best_f1=0.9351851851851852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09586045145988464
step: 10, loss: 0.10489052534103394
step: 20, loss: 0.06287495791912079
step: 30, loss: 0.010168460197746754
step: 40, loss: 0.0034480353351682425
step: 50, loss: 0.10977970063686371
step: 60, loss: 0.1560390591621399
step: 70, loss: 0.06750358641147614
step: 80, loss: 0.02043640986084938
step: 90, loss: 0.009768929332494736
step: 100, loss: 0.37158551812171936
step: 110, loss: 0.07488419860601425
step: 120, loss: 0.07067988067865372
step: 130, loss: 0.015060029923915863
step: 140, loss: 0.0163356214761734
step: 150, loss: 0.1151900365948677
step: 160, loss: 0.04425628110766411
step: 170, loss: 0.13181409239768982
step: 180, loss: 0.019628720358014107
step: 190, loss: 0.022430740296840668
step: 200, loss: 0.008728829212486744
step: 210, loss: 0.06811235845088959
step: 220, loss: 0.08389924466609955
step: 230, loss: 0.0786760151386261
step: 240, loss: 0.11304957419633865
step: 250, loss: 0.08198504149913788
step: 260, loss: 0.012373422272503376
step: 270, loss: 0.29390397667884827
step: 280, loss: 0.20682261884212494
step: 290, loss: 0.10837748646736145
step: 300, loss: 0.029667170718312263
step: 310, loss: 0.035269055515527725
step: 320, loss: 0.1202320083975792
```

```
step: 330, loss: 0.025512393563985825
step: 340, loss: 0.004973718896508217
step: 350, loss: 0.05365399643778801
step: 360, loss: 0.010871653445065022
step: 370, loss: 0.006666928995400667
step: 380, loss: 0.0750090554356575
step: 390, loss: 0.0162032637745142
step: 400, loss: 0.036721616983413696
step: 410, loss: 0.0020285563077777624
step: 420, loss: 0.06634407490491867
step: 430, loss: 0.012458359822630882
step: 440, loss: 0.014085536822676659
step: 450, loss: 0.02213946357369423
step: 460, loss: 0.16766850650310516
step: 470, loss: 0.0696747899055481
step: 480, loss: 0.21196258068084717
step: 490, loss: 0.041483741253614426
step: 500, loss: 0.013399526476860046
step: 510, loss: 0.03334760665893555
step: 520, loss: 0.022204777225852013
step: 530, loss: 0.16977520287036896
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9398040130657956, f1=0.9428704133766836,
best_f1=0.9428704133766836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036476380191743374
step: 10, loss: 0.08021976053714752
step: 20, loss: 0.13495640456676483
step: 30, loss: 0.1842544674873352
step: 40, loss: 0.005890390370041132
step: 50, loss: 0.0075763110007142067
step: 60, loss: 0.0018060167785733938
step: 70, loss: 0.056855082511901855
step: 80, loss: 0.0007389477686956525
step: 90, loss: 0.041935138404369354
step: 100, loss: 0.022613922134041786
step: 110, loss: 0.0071860747411847115
step: 120, loss: 0.014438568614423275
step: 130, loss: 0.013970789499958086
step: 140, loss: 0.014987356029450893
```

```
step: 150, loss: 0.0026871152222156525
step: 160, loss: 0.0005651554092764854
step: 170, loss: 0.008490710519254208
step: 180, loss: 0.03822127357125282
step: 190, loss: 0.04650430008769035
step: 200, loss: 0.01331986952573061
step: 210, loss: 0.06823360174894333
step: 220, loss: 0.02326779253780842
step: 230, loss: 0.04046949744224548
step: 240, loss: 0.0029094477649778128
step: 250, loss: 0.004077276214957237
step: 260, loss: 0.004374518524855375
step: 270, loss: 0.0008858832297846675
step: 280, loss: 0.06445135176181793
step: 290, loss: 0.06153123080730438
step: 300, loss: 0.058666545897722244
step: 310, loss: 0.1961083710193634
step: 320, loss: 0.07865173369646072
step: 330, loss: 0.0057181078009307384
step: 340, loss: 0.0021088523790240288
step: 350, loss: 0.008072114549577236
step: 360, loss: 0.0068764160387218
step: 370, loss: 0.0012703993124887347
step: 380, loss: 0.007798677310347557
step: 390, loss: 0.03974176570773125
step: 400, loss: 0.029329027980566025
step: 410, loss: 0.009522907435894012
step: 420, loss: 0.06607815623283386
step: 430, loss: 0.055674389004707336
step: 440, loss: 0.016174783930182457
step: 450, loss: 0.07888554036617279
step: 460, loss: 0.03269507363438606
step: 470, loss: 0.014809099026024342
step: 480, loss: 0.002967524342238903
step: 490, loss: 0.002211355371400714
step: 500, loss: 0.16306515038013458
step: 510, loss: 0.006656920071691275
step: 520, loss: 0.005188826471567154
step: 530, loss: 0.09875334054231644
/usr/local/lib/python3.7/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=0.9368273280299485, f1=0.9371800837598881,
best_f1=0.9428704133766836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002355733886361122
```

```
step: 10, loss: 0.007011493667960167
step: 20, loss: 0.024918096140027046
step: 30, loss: 0.005859294906258583
step: 40, loss: 0.003300377866253257
step: 50, loss: 0.0018707637209445238
step: 60, loss: 0.0003291663888376206
step: 70, loss: 0.000784909469075501
step: 80, loss: 0.0018724241526797414
step: 90, loss: 0.07506681233644485
step: 100, loss: 0.10058069974184036
step: 110, loss: 0.0014134965604171157
step: 120, loss: 0.000782400369644165
step: 130, loss: 0.09338818490505219
step: 140, loss: 0.007482773624360561
step: 150, loss: 0.004221681505441666
step: 160, loss: 0.014905139803886414
step: 170, loss: 0.005153411999344826
step: 180, loss: 0.07892577350139618
step: 190, loss: 0.009514538571238518
step: 200, loss: 0.0008808559505268931
step: 210, loss: 0.03620457649230957
step: 220, loss: 0.007165322080254555
step: 230, loss: 0.15777352452278137
step: 240, loss: 0.0009780635591596365
step: 250, loss: 0.01140500232577324
step: 260, loss: 0.08224070072174072
step: 270, loss: 0.02498524822294712
step: 280, loss: 0.05368060991168022
step: 290, loss: 0.11885342001914978
step: 300, loss: 0.00245024310424239
step: 310, loss: 0.0291998740285635
step: 320, loss: 0.007000345271080732
step: 330, loss: 0.003292855340987444
step: 340, loss: 0.019130107015371323
step: 350, loss: 0.0030290037393569946
step: 360, loss: 0.012696603313088417
step: 370, loss: 0.041172005236148834
step: 380, loss: 0.00281738699413836
step: 390, loss: 0.03019952028989792
step: 400, loss: 0.01001804880797863
step: 410, loss: 0.005758496001362801
step: 420, loss: 0.01022802758961916
step: 430, loss: 0.0022330195643007755
step: 440, loss: 0.07684449851512909
step: 450, loss: 0.004714405629783869
step: 460, loss: 0.00040152997826226056
step: 470, loss: 0.002016239333897829
step: 480, loss: 0.00922127440571785
```

```
step: 490, loss: 0.029053281992673874
step: 500, loss: 0.005686081014573574
step: 510, loss: 0.01836494728922844
step: 520, loss: 0.014084259048104286
step: 530, loss: 0.005615738686174154
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9419537517697026, f1=0.935285781766651,
best_f1=0.935285781766651
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00270420522429049
step: 10, loss: 0.037317629903554916
step: 20, loss: 0.02552337571978569
step: 30, loss: 0.011319209821522236
step: 40, loss: 0.02638404443860054
step: 50, loss: 0.040099117904901505
step: 60, loss: 0.005524720996618271
step: 70, loss: 0.0011926166480407119
step: 80, loss: 0.00071836466666817367
step: 90, loss: 0.0032066311687231064
step: 100, loss: 0.003912649117410183
step: 110, loss: 0.0042923432774841785
step: 120, loss: 0.004756685812026262
step: 130, loss: 0.0008266504155471921
step: 140, loss: 0.0010271637002006173
step: 150, loss: 0.0006148903048597276
step: 160, loss: 0.0015687921550124884
step: 170, loss: 0.009891319088637829
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 180, loss: 0.023120664060115814
step: 190, loss: 0.00048524473095312715
step: 200, loss: 0.0020254370756447315
step: 210, loss: 0.0009900500299409032
step: 220, loss: 0.008090460672974586
step: 230, loss: 0.0022617261856794357
step: 240, loss: 0.005226743873208761
step: 250, loss: 0.00023416909971274436
step: 260, loss: 0.005204102024435997
step: 270, loss: 0.00450331112369895
step: 280, loss: 0.06195225194096565
step: 290, loss: 0.000659215496852994
```

```
step: 300, loss: 0.08168371021747589
step: 310, loss: 0.00034402761957608163
step: 320, loss: 0.023977458477020264
step: 330, loss: 0.007005791179835796
step: 340, loss: 0.027721568942070007
step: 350, loss: 0.0024066052865236998
step: 360, loss: 0.0017768157413229346
step: 370, loss: 0.0025000711902976036
step: 380, loss: 0.07536819577217102
step: 390, loss: 0.012974248267710209
step: 400, loss: 0.04085070267319679
step: 410, loss: 0.00033400801476091146
step: 420, loss: 0.010796518065035343
step: 430, loss: 0.010811042040586472
step: 440, loss: 0.001905796118080616
step: 450, loss: 0.0042934725061059
step: 460, loss: 0.01063312217593193
step: 470, loss: 0.005928277038037777
step: 480, loss: 0.005265348125249147
step: 490, loss: 0.001302745658904314
step: 500, loss: 0.0025494189467281103
step: 510, loss: 0.014799805358052254
step: 520, loss: 0.0009418059489689767
step: 530, loss: 0.053177665919065475
/usr/local/lib/python3.7/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=0.9402985074626865, f1=0.9408502772643254,
best_f1=0.935285781766651
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007374018314294517
step: 10, loss: 0.004758214112371206
step: 20, loss: 0.0005724749644286931
step: 30, loss: 0.0007814339478500187
step: 40, loss: 0.00037864860496483743
step: 50, loss: 0.0008794068126007915
step: 60, loss: 0.0020633679814636707
step: 70, loss: 0.07178308069705963
step: 80, loss: 0.00016373897960875183
step: 90, loss: 0.0005366806290112436
step: 100, loss: 0.04561389610171318
step: 110, loss: 0.006406164728105068
step: 120, loss: 0.0004515088221523911
step: 130, loss: 0.018353689461946487
step: 140, loss: 0.00025507318787276745
step: 150, loss: 0.008367582224309444
```

```
step: 160, loss: 0.00010631881741574034
step: 170, loss: 0.0004655159136746079
step: 180, loss: 0.0007632015040144324
step: 190, loss: 0.003102088114246726
step: 200, loss: 7.010102126514539e-05
step: 210, loss: 0.004125235602259636
step: 220, loss: 0.0013559602666646242
step: 230, loss: 0.00012262076779734343
step: 240, loss: 0.014469554647803307
step: 250, loss: 0.00021578943415079266
step: 260, loss: 0.002938257297500968
step: 270, loss: 0.0037000360898673534
step: 280, loss: 0.0005950563936494291
step: 290, loss: 0.0037747370079159737
step: 300, loss: 0.002353595569729805
step: 310, loss: 0.012237511575222015
step: 320, loss: 0.0875689908862114
step: 330, loss: 0.003592574270442128
step: 340, loss: 0.002007271395996213
step: 350, loss: 0.051883939653635025
step: 360, loss: 9.068202780326828e-05
step: 370, loss: 0.0377490371465683
step: 380, loss: 0.158451646566391
step: 390, loss: 0.0328984335064888
step: 400, loss: 9.854751988314092e-05
step: 410, loss: 0.0015076822601258755
step: 420, loss: 0.010151054710149765
step: 430, loss: 0.00039998613647185266
step: 440, loss: 0.002541513182222843
step: 450, loss: 0.004511762876063585
step: 460, loss: 0.000578309700358659
step: 470, loss: 0.2430175393819809
step: 480, loss: 0.01858758181333542
step: 490, loss: 0.001916856854222715
step: 500, loss: 0.004072765354067087
step: 510, loss: 0.0008725534426048398
step: 520, loss: 0.010664230212569237
step: 530, loss: 0.0003446415939833969
/usr/local/lib/python3.7/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=0.9410125406409661, f1=0.9380530973451328,
best_f1=0.935285781766651
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023836090695112944
step: 10, loss: 0.007174766156822443
```

```
step: 20, loss: 0.005694999825209379
step: 30, loss: 0.006310997065156698
step: 40, loss: 0.0003384824376553297
step: 50, loss: 0.0029654621612280607
step: 60, loss: 0.0006733203190378845
step: 70, loss: 0.18388515710830688
step: 80, loss: 0.02914535254240036
step: 90, loss: 0.0007579989614896476
step: 100, loss: 0.009807324968278408
step: 110, loss: 0.004542254842817783
step: 120, loss: 0.0003266730927862227
step: 130, loss: 0.00018434427329339087
step: 140, loss: 0.0007791620446369052
step: 150, loss: 0.00016071890422608703
step: 160, loss: 0.0007831210386939347
step: 170, loss: 0.0007317448034882545
step: 180, loss: 0.0002882335102185607
step: 190, loss: 0.0001106961935875006
step: 200, loss: 0.0007078216876834631
step: 210, loss: 0.0024463615845888853
step: 220, loss: 0.0006020930013619363
step: 230, loss: 0.0006912382086738944
step: 240, loss: 0.002459055045619607
step: 250, loss: 0.001266407547518611
step: 260, loss: 0.0003149424446746707
step: 270, loss: 7.314461981877685e-05
step: 280, loss: 0.0032707061618566513
step: 290, loss: 0.00424790196120739
step: 300, loss: 0.002247256226837635
step: 310, loss: 0.00027127141947858036
step: 320, loss: 0.02942344732582569
step: 330, loss: 0.07209809124469757
step: 340, loss: 0.001683372538536787
step: 350, loss: 0.012150226160883904
step: 360, loss: 0.0033731642179191113
step: 370, loss: 0.0003231470473110676
step: 380, loss: 0.002070294227451086
step: 390, loss: 0.00016805923951324075
step: 400, loss: 0.00013724082964472473
step: 410, loss: 0.0018516259733587503
step: 420, loss: 0.0003462271997705102
step: 430, loss: 0.0003295032656751573
step: 440, loss: 0.0012967042857781053
step: 450, loss: 0.000240959067014046
step: 460, loss: 0.003521872917190194
step: 470, loss: 0.007496486883610487
step: 480, loss: 0.016371872276067734
step: 490, loss: 0.00041950360173359513
```

```
step: 500, loss: 9.439400309929624e-05
step: 510, loss: 0.001385182491503656
step: 520, loss: 0.0016211378388106823
step: 530, loss: 7.417969027301297e-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=0.9433255269320844, f1=0.9429906542056075,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0022236425429582596
step: 10, loss: 0.01207207702100277
step: 20, loss: 0.016758309677243233
step: 30, loss: 0.01566682569682598
step: 40, loss: 7.835430733393878e-05
step: 50, loss: 0.00057904602726921444
step: 60, loss: 0.005935593042522669
step: 70, loss: 0.0020394213497638702
step: 80, loss: 0.069207176566123396
step: 90, loss: 0.00017185629985760897
step: 100, loss: 0.004106159321963787
step: 110, loss: 0.0013137500500306487
step: 120, loss: 0.012941054068505764
step: 130, loss: 0.016187358647584915
step: 140, loss: 3.879614087054506e-05
step: 150, loss: 0.0009317845688201487
step: 160, loss: 0.0015073977410793304
step: 170, loss: 0.00035692082019522786
step: 180, loss: 0.0004027241375297308
step: 190, loss: 0.001095243962481618
step: 200, loss: 0.000801873451564461
step: 210, loss: 0.000325339671690017
step: 220, loss: 0.0016053832368925214
step: 230, loss: 0.02434062957763672
step: 240, loss: 0.00038301688618957996
step: 250, loss: 0.0001421943452442065
step: 260, loss: 0.024138957262039185
step: 270, loss: 6.248700810829177e-05
step: 280, loss: 0.0001576724898768589
step: 290, loss: 0.007163751404732466
step: 300, loss: 3.60775702574756e-05
step: 310, loss: 0.07006681710481644
```

```
step: 320, loss: 0.00015405012527480721
step: 330, loss: 0.02320726402103901
step: 340, loss: 0.0036429588217288256
step: 350, loss: 0.012591082602739334
step: 360, loss: 0.0001596658694325015
step: 370, loss: 0.0007352267275564373
step: 380, loss: 0.010200917720794678
step: 390, loss: 0.00016332116501871496
step: 400, loss: 0.05865398794412613
step: 410, loss: 0.00022237995290197432
step: 420, loss: 0.0001510522997705266
step: 430, loss: 0.00015935210103634745
step: 440, loss: 0.00027527709607966244
step: 450, loss: 0.0008520736591890454
step: 460, loss: 0.0023207750637084246
step: 470, loss: 0.0011631298111751676
step: 480, loss: 0.0004995877388864756
step: 490, loss: 0.0026916104834526777
step: 500, loss: 0.003529266221448779
step: 510, loss: 8.504636207362637e-05
step: 520, loss: 0.0001015167435980402
step: 530, loss: 0.00017620147264096886
/usr/local/lib/python3.7/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=0.9425393883225208, f1=0.9423165666820489,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027514738962054253
step: 10, loss: 0.00027018555555642396
step: 20, loss: 8.793111919658259e-05
step: 30, loss: 0.0016699290135875344
step: 40, loss: 0.0002199996670242399
step: 50, loss: 0.0003371519851498306
step: 60, loss: 0.00017470514285378158
step: 70, loss: 0.12379005551338196
step: 80, loss: 8.474812057102099e-05
step: 90, loss: 0.010162698104977608
step: 100, loss: 0.0001376974250888452
step: 110, loss: 0.005061053670942783
step: 120, loss: 6.070656672818586e-05
step: 130, loss: 6.908985233167186e-05
step: 140, loss: 0.0004651947820093483
step: 150, loss: 8.495986548950896e-05
step: 160, loss: 0.000871148775331676
step: 170, loss: 7.311891386052594e-05
```

```
step: 180, loss: 6.667213165201247e-05
step: 190, loss: 0.0010278564877808094
step: 200, loss: 0.0036528324708342552
step: 210, loss: 0.00028222458786331117
step: 220, loss: 0.026079166680574417
step: 230, loss: 0.0006765893194824457
step: 240, loss: 0.02092113345861435
step: 250, loss: 0.0005743344081565738
step: 260, loss: 0.0005550503265112638
step: 270, loss: 0.001155270030722022
step: 280, loss: 6.173941801534966e-05
step: 290, loss: 4.387411900097504e-05
step: 300, loss: 0.00024429973564110696
step: 310, loss: 3.439070860622451e-05
step: 320, loss: 0.00011107102181995288
step: 330, loss: 0.019109342247247696
step: 340, loss: 4.9827442126115784e-05
step: 350, loss: 6.975099677219987e-05
step: 360, loss: 0.026254262775182724
step: 370, loss: 0.000308683724142611
step: 380, loss: 0.00025580829242244363
step: 390, loss: 0.00023314272402785718
step: 400, loss: 0.025993352755904198
step: 410, loss: 0.00021972678950987756
step: 420, loss: 0.00012197127944091335
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 430, loss: 4.681285281549208e-05
step: 440, loss: 0.00022885703947395086
step: 450, loss: 0.00015542563050985336
step: 460, loss: 0.000941131729632616
step: 470, loss: 0.003254452720284462
step: 480, loss: 0.0006233112653717399
step: 490, loss: 0.000392250920413062
step: 500, loss: 0.004196300636976957
step: 510, loss: 0.000987838488072157
step: 520, loss: 0.0016545733669772744
step: 530, loss: 0.0004586548893712461
/usr/local/lib/python3.7/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=0.9383177570093458, f1=0.9374416433239963,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.054306358098983765
step: 10, loss: 0.0007734128157608211
step: 20, loss: 0.014544122852385044
```

```
step: 30, loss: 0.00032902052043937147
step: 40, loss: 0.0006244418909773231
step: 50, loss: 0.00046803514123894274
step: 60, loss: 0.0007676269160583615
step: 70, loss: 0.0009440130670554936
step: 80, loss: 0.0059964414685964584
step: 90, loss: 0.00022020080359652638
step: 100, loss: 0.0004898237530142069
step: 110, loss: 0.001830501714721322
step: 120, loss: 0.00014932575868442655
step: 130, loss: 0.00043495246791280806
step: 140, loss: 0.0013427651720121503
step: 150, loss: 0.00038099047378636897
step: 160, loss: 0.0005014828057028353
step: 170, loss: 0.002022698987275362
step: 180, loss: 0.007964488118886948
step: 190, loss: 0.00028999013011343777
step: 200, loss: 0.0003592920256778598
step: 210, loss: 0.005900566000491381
step: 220, loss: 0.0003605254751164466
step: 230, loss: 0.0004964323597960174
step: 240, loss: 4.023174915346317e-05
step: 250, loss: 0.008155470713973045
step: 260, loss: 0.00548802362754941
step: 270, loss: 0.0011117026442661881
step: 280, loss: 0.006892075762152672
step: 290, loss: 0.0001664298470132053
step: 300, loss: 0.00025592290330678225
step: 310, loss: 5.6097531341947615e-05
step: 320, loss: 0.00011629833898041397
step: 330, loss: 0.0005269249668344855
step: 340, loss: 8.741229248698801e-05
step: 350, loss: 3.220798680558801e-05
step: 360, loss: 0.0008601978188380599
step: 370, loss: 0.0006575300358235836
step: 380, loss: 0.00024946965277194977
step: 390, loss: 2.991675319208298e-05
step: 400, loss: 0.00011184895993210375
step: 410, loss: 0.0003262867103330791
step: 420, loss: 0.005165507085621357
step: 430, loss: 4.526834163698368e-05
step: 440, loss: 0.001016973052173853
step: 450, loss: 0.029523510485887527
step: 460, loss: 0.00014448401634581387
step: 470, loss: 4.5739583583781496e-05
step: 480, loss: 0.001362502807751298
step: 490, loss: 0.042695775628089905
step: 500, loss: 0.009717365726828575
```

```
step: 510, loss: 0.0003245431580580771
step: 520, loss: 0.002711015520617366
step: 530, loss: 0.00015080069715622813
/usr/local/lib/python3.7/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=0.9383624655013799, f1=0.9412844036697248,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.041976310312747955
step: 10, loss: 0.014590836130082607
step: 20, loss: 0.0008716621086932719
step: 30, loss: 0.0005828170687891543
step: 40, loss: 0.029555613175034523
step: 50, loss: 0.0015137989539653063
step: 60, loss: 8.818024070933461e-05
step: 70, loss: 0.00016543730453122407
step: 80, loss: 0.000561877153813839
step: 90, loss: 0.00017007642600219697
step: 100, loss: 4.877114042756148e-05
step: 110, loss: 0.001334781525656581
step: 120, loss: 0.00016467207751702517
step: 130, loss: 5.65799382457044e-05
step: 140, loss: 0.00021075105178169906
step: 150, loss: 0.0001716526603559032
step: 160, loss: 0.0002635618729982525
step: 170, loss: 0.002457362599670887
step: 180, loss: 0.0001282793964492157
step: 190, loss: 0.006012882571667433
step: 200, loss: 0.0019232393242418766
step: 210, loss: 0.00010107782145496458
step: 220, loss: 0.0002079820988001302
step: 230, loss: 0.00017110705084633082
step: 240, loss: 2.5383302272530273e-05
step: 250, loss: 0.0011425931006669998
step: 260, loss: 0.0015832330100238323
step: 270, loss: 0.00046788417967036366
step: 280, loss: 0.00032342603662982583
step: 290, loss: 0.00044841745775193
step: 300, loss: 0.028547832742333412
step: 310, loss: 9.879495337372646e-05
step: 320, loss: 0.02329058200120926
step: 330, loss: 0.0010230459738522768
step: 340, loss: 3.206205292372033e-05
step: 350, loss: 0.017082899808883667
step: 360, loss: 0.003396068001165986
```

```
step: 370, loss: 2.1300533262547106e-05
step: 380, loss: 6.0077709349570796e-05
step: 390, loss: 0.0027130343951284885
step: 400, loss: 2.7092428354080766e-05
step: 410, loss: 0.005627896171063185
step: 420, loss: 0.00011685677600326017
step: 430, loss: 0.00029306215583346784
step: 440, loss: 5.0913771701743826e-05
step: 450, loss: 3.517616642056964e-05
step: 460, loss: 0.0018211085116490722
step: 470, loss: 0.05217871442437172
step: 480, loss: 0.0001882184442365542
step: 490, loss: 0.00048269116086885333
step: 500, loss: 5.300427437759936e-05
step: 510, loss: 3.797355748247355e-05
step: 520, loss: 3.4348729968769476e-05
step: 530, loss: 5.3900825150776654e-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=0.9401392111368909, f1=0.9403606102635228,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.4908789024921134e-05
step: 10, loss: 0.00033262796932831407
step: 20, loss: 2.7859552574227564e-05
step: 30, loss: 0.00029317432199960481
step: 40, loss: 4.183297642157413e-05
step: 50, loss: 0.005398192442953587
step: 60, loss: 4.5035438233753666e-05
step: 70, loss: 0.00087949715089905292
step: 80, loss: 0.002633965341374278
step: 90, loss: 0.0009178972686640918
step: 100, loss: 0.00013153473264537752
step: 110, loss: 2.5852508770185523e-05
step: 120, loss: 0.000697753974236548
step: 130, loss: 3.9974096580408514e-05
step: 140, loss: 0.00016248581232503057
step: 150, loss: 0.00494769960641861
step: 160, loss: 0.0012990983668714762
step: 170, loss: 0.0002885850553866476
step: 180, loss: 0.0019635779317468405
step: 190, loss: 1.909551974677015e-05
step: 200, loss: 6.720512465108186e-05
step: 210, loss: 0.00015154338325373828
step: 220, loss: 2.608748218335677e-05
```

```
step: 230, loss: 0.01590033620595932
step: 240, loss: 0.0004806483630090952
step: 250, loss: 3.367035969858989e-05
step: 260, loss: 0.012553545646369457
step: 270, loss: 2.2418111257138662e-05
step: 280, loss: 0.00026826324756257236
step: 290, loss: 0.0024832230992615223
step: 300, loss: 0.0002772305451799184
step: 310, loss: 5.3032497817184776e-05
step: 320, loss: 0.00010465586819918826
step: 330, loss: 2.6858271667151712e-05
step: 340, loss: 0.0015172804705798626
step: 350, loss: 0.02008660137653351
step: 360, loss: 7.671215280424803e-05
step: 370, loss: 2.8217707949806936e-05
step: 380, loss: 0.00010189929162152112
step: 390, loss: 0.006353600416332483
step: 400, loss: 9.188734838971868e-05
step: 410, loss: 0.0003418749547563493
step: 420, loss: 0.0010993147734552622
step: 430, loss: 2.725712511164602e-05
step: 440, loss: 0.0008754768641665578
step: 450, loss: 7.255262607941404e-05
step: 460, loss: 3.724334601429291e-05
step: 470, loss: 2.878391205740627e-05
step: 480, loss: 0.00011087262100772932
step: 490, loss: 0.0006563634378835559
step: 500, loss: 0.0033839792013168335
step: 510, loss: 2.493282408977393e-05
step: 520, loss: 0.024981459602713585
step: 530, loss: 0.0020116502419114113
/usr/local/lib/python3.7/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=0.9406307977736549, f1=0.943344081068632,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.7405460059526376e-05
step: 10, loss: 1.5273513781721704e-05
step: 20, loss: 0.017331033945083618
step: 30, loss: 0.001290869782678783
step: 40, loss: 0.0005281983176246285
step: 50, loss: 2.0667450371547602e-05
step: 60, loss: 7.462102075805888e-05
step: 70, loss: 2.012736877077259e-05
step: 80, loss: 1.7486216165707447e-05
```

```
step: 90, loss: 0.0013504582457244396
step: 100, loss: 0.0002875060308724642
step: 110, loss: 5.5117827287176624e-05
step: 120, loss: 0.00010232964996248484
step: 130, loss: 0.002075823023915291
step: 140, loss: 5.5317730584647506e-05
step: 150, loss: 2.742856668191962e-05
step: 160, loss: 0.0002898165548685938
step: 170, loss: 0.0005980214918963611
step: 180, loss: 3.507059227558784e-05
step: 190, loss: 1.919984060805291e-05
step: 200, loss: 0.001533050206489861
step: 210, loss: 0.00030303365201689303
step: 220, loss: 2.0555731680360623e-05
step: 230, loss: 0.02458416111767292
step: 240, loss: 3.652398299891502e-05
step: 250, loss: 0.07286665588617325
step: 260, loss: 0.0002800219226628542
step: 270, loss: 0.00011526331218192354
step: 280, loss: 2.8649874366237782e-05
step: 290, loss: 2.411686728009954e-05
step: 300, loss: 0.00016089691780507565
step: 310, loss: 6.748683517798781e-05
step: 320, loss: 6.214059249032289e-05
step: 330, loss: 6.727795698679984e-05
step: 340, loss: 7.384059426840395e-05
step: 350, loss: 0.03108929842710495
step: 360, loss: 2.6172565412707627e-05
step: 370, loss: 4.950923903379589e-05
step: 380, loss: 0.0012960339663550258
step: 390, loss: 3.533874041750096e-05
step: 400, loss: 8.921040716813877e-05
step: 410, loss: 0.00023278470325749367
step: 420, loss: 7.353939872700721e-05
step: 430, loss: 8.550426719011739e-05
step: 440, loss: 7.434860890498385e-05
step: 450, loss: 0.0002618737635202706
step: 460, loss: 2.8711661798297428e-05
step: 470, loss: 0.019514426589012146
step: 480, loss: 0.00017643866885919124
step: 490, loss: 5.022957338951528e-05
step: 500, loss: 9.897129348246381e-05
step: 510, loss: 0.00017195608234032989
step: 520, loss: 2.0309456886025146e-05
step: 530, loss: 0.0004600288812071085
/usr/local/lib/python3.7/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=0.9388523047977423, f1=0.9427107591988821,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.375640416285023e-05
step: 10, loss: 3.92618530895561e-05
step: 20, loss: 1.8439752238919027e-05
step: 30, loss: 2.4727902200538665e-05
step: 40, loss: 0.0001759421284077689
step: 50, loss: 0.00034585242974571884
step: 60, loss: 5.461878754431382e-05
step: 70, loss: 2.017558836087119e-05
step: 80, loss: 0.00014007290883455426
step: 90, loss: 2.7440994017524645e-05
step: 100, loss: 3.2766190997790545e-05
step: 110, loss: 1.5329373127315193e-05
step: 120, loss: 0.00016101720393635333
step: 130, loss: 9.826155292103067e-05
step: 140, loss: 0.01959828846156597
step: 150, loss: 3.3259497286053374e-05
step: 160, loss: 0.001632552593946457
step: 170, loss: 0.0004190715553704649
step: 180, loss: 3.2951862522168085e-05
step: 190, loss: 5.295359005685896e-05
step: 200, loss: 3.5353979910723865e-05
step: 210, loss: 0.0002814435865730047
step: 220, loss: 8.363567030755803e-05
step: 230, loss: 0.001928993733599782
step: 240, loss: 0.015476824715733528
step: 250, loss: 2.162474993383512e-05
step: 260, loss: 1.2226309991092421e-05
step: 270, loss: 0.0050956252962350845
step: 280, loss: 4.053532757097855e-05
step: 290, loss: 2.7543073883862235e-05
step: 300, loss: 0.00012293607869651169
step: 310, loss: 2.9195070965215564e-05
step: 320, loss: 3.3572709071449935e-05
step: 330, loss: 0.000942815444432199
step: 340, loss: 2.3286251234821975e-05
step: 350, loss: 0.00021595502039417624
step: 360, loss: 0.004363791085779667
step: 370, loss: 2.1702922822441906e-05
step: 380, loss: 2.2280040866462514e-05
step: 390, loss: 2.128163760062307e-05
step: 400, loss: 0.0007042839424684644
step: 410, loss: 1.5698189599788748e-05
step: 420, loss: 1.319857983617112e-05
```

```
step: 430, loss: 2.338973899895791e-05
step: 440, loss: 1.66517111210851e-05
step: 450, loss: 2.3755175789119676e-05
step: 460, loss: 0.0003354123036842793
step: 470, loss: 2.258937638544012e-05
step: 480, loss: 1.6092964870040305e-05
step: 490, loss: 6.670237780781463e-05
step: 500, loss: 0.00010473492147866637
step: 510, loss: 8.854826592141762e-05
step: 520, loss: 2.0049175873282366e-05
step: 530, loss: 8.799534407444298e-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=0.9399624765478424, f1=0.9447282861124013,
best_f1=0.9429906542056075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014191675931215286
step: 10, loss: 4.215261651552282e-05
step: 20, loss: 2.2015867216396146e-05
step: 30, loss: 6.21875369688496e-05
step: 40, loss: 1.737808997859247e-05
step: 50, loss: 0.0013043474173173308
step: 60, loss: 1.8640505004441366e-05
step: 70, loss: 6.26902183284983e-05
step: 80, loss: 5.90572708460968e-05
step: 90, loss: 1.3347586900636088e-05
step: 100, loss: 1.3720108654524665e-05
step: 110, loss: 0.01436868216842413
step: 120, loss: 0.0017349018016830087
step: 130, loss: 1.8175272998632863e-05
step: 140, loss: 1.5809759133844636e-05
step: 150, loss: 0.00022790073126088828
step: 160, loss: 0.008666245266795158
step: 170, loss: 9.207645052811131e-05
step: 180, loss: 4.5525655878009275e-05
step: 190, loss: 7.16191716492176e-05
step: 200, loss: 6.163079524412751e-05
step: 210, loss: 2.185544144595042e-05
step: 220, loss: 1.5567658920190297e-05
step: 230, loss: 0.013928010128438473
step: 240, loss: 4.857184467255138e-05
step: 250, loss: 4.084057582076639e-05
step: 260, loss: 8.807343692751601e-05
step: 270, loss: 0.008502203971147537
step: 280, loss: 5.942075949860737e-05
```

```
step: 290, loss: 0.0024715824984014034
step: 300, loss: 0.0021798633970320225
step: 310, loss: 4.052659642184153e-05
step: 320, loss: 5.635638808598742e-05
step: 330, loss: 2.259310713270679e-05
step: 340, loss: 8.349926065420732e-05
step: 350, loss: 1.8242413716507144e-05
step: 360, loss: 4.8769892600830644e-05
step: 370, loss: 1.4599246242141817e-05
step: 380, loss: 1.4383199413714465e-05
step: 390, loss: 1.8283377357874997e-05
step: 400, loss: 1.2684510693361517e-05
step: 410, loss: 0.0007525432156398892
step: 420, loss: 2.164721445296891e-05
step: 430, loss: 2.760198913165368e-05
step: 440, loss: 6.350650801323354e-05
step: 450, loss: 0.0013034981675446033
step: 460, loss: 2.4584373022662476e-05
step: 470, loss: 0.000999510521069169
step: 480, loss: 1.8961340174428187e-05
step: 490, loss: 0.000193214655155316
step: 500, loss: 4.8990052164299414e-05
step: 510, loss: 2.6325220460421406e-05
step: 520, loss: 3.845048922812566e-05
step: 530, loss: 2.7792828404926695e-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=0.9392369288742346, f1=0.9445221445221446,
best_f1=0.9429906542056075
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
```

```
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:13, 427.39it/s]
load_f1 = 0.9398280802292264
real_f1 = 0.9375298044825942
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 435.29it/s]
```

1.1.4   Amazon-Google - Running the matcher

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

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

     # Step 3: run the matcher
     !CUDA_VISIBLE_DEVICES=0 python matcher.py \
       --task Structured/Amazon-Google \
```

```
    --input_path input/input_small.jsonl \
    --output_path output/output_small.jsonl \
    --lm distilbert \
    --use_gpu \
    --fp16 \
    --checkpoint_path checkpoints/
```

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

Defaults for this optimization level are:
enabled   : True
opt_level   : O2
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights   : True
loss_scale   : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled   : True
opt_level   : O2
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights   : True
loss_scale   : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
  "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.831695020198822
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06263329088687897
step: 20, loss: 0.38546743988990784

```
step: 30, loss: 0.36778202652931213
step: 40, loss: 0.4890604615211487
step: 50, loss: 0.2900713384151459
step: 60, loss: 0.35378360748291016
step: 70, loss: 0.19338767230510712
step: 80, loss: 0.347426176071167
step: 90, loss: 0.37464940547943115
step: 100, loss: 0.08913303911685944
step: 110, loss: 0.3140937387943268
step: 120, loss: 0.2429915815591812
step: 130, loss: 0.1726216822862625
step: 140, loss: 0.24124203622341156
step: 150, loss: 0.22977636754512787
step: 160, loss: 0.14247897267341614
step: 170, loss: 0.08491650223731995
step: 180, loss: 0.15085236728191376
step: 190, loss: 0.17812439799308777
step: 200, loss: 0.13160856068134308
step: 210, loss: 0.35851791501045227
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6309278350515464, f1=0.6945054945054946,
best_f1=0.6945054945054946
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03897567465901375
step: 10, loss: 0.07914516329765329
step: 20, loss: 0.179937943816185
step: 30, loss: 0.15989887714385986
step: 40, loss: 0.1065308004617691
step: 50, loss: 0.20913589000701904
step: 60, loss: 0.02763730101287365
step: 70, loss: 0.1548272967338562
step: 80, loss: 0.15758484601974487
step: 90, loss: 0.07010933011770248
step: 100, loss: 0.04278067871928215
step: 110, loss: 0.1326320618391037
step: 120, loss: 0.16279534995555878
step: 130, loss: 0.16631805896759033
step: 140, loss: 0.11612950265407562
step: 150, loss: 0.18336616456508636
step: 160, loss: 0.14060848951339722
```

```
step: 170, loss: 0.1513778269290924
step: 180, loss: 0.21707883477210999
step: 190, loss: 0.08038109540939331
step: 200, loss: 0.32893800735473633
step: 210, loss: 0.30922019481658936
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.676056338028169, f1=0.7044534412955465,
best_f1=0.7044534412955465
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19198822975158691
step: 10, loss: 0.1786891222000122
step: 20, loss: 0.2569355070590973
step: 30, loss: 0.08724749833345413
step: 40, loss: 0.1477440893650055
step: 50, loss: 0.1451389491558075
step: 60, loss: 0.23491840064525604
step: 70, loss: 0.1059672087430954
step: 80, loss: 0.043482303619384766
step: 90, loss: 0.1934654861688614
step: 100, loss: 0.020165108144283295
step: 110, loss: 0.2754482328891754
step: 120, loss: 0.19364696741104126
step: 130, loss: 0.15010139346122742
step: 140, loss: 0.21706189215183258
step: 150, loss: 0.14704053103923798
step: 160, loss: 0.08638278394937515
step: 170, loss: 0.1562957912683487
step: 180, loss: 0.04903717339038849
step: 190, loss: 0.11655684560537338
step: 200, loss: 0.1088971495628357
step: 210, loss: 0.13347619771957397
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6945454545454546, f1=0.7065026362038664,
best_f1=0.7065026362038664
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21641229093074799
step: 10, loss: 0.015067433007061481
step: 20, loss: 0.09320612251758575
step: 30, loss: 0.052398160099983215
step: 40, loss: 0.08496464788913727
step: 50, loss: 0.047485820949077606
step: 60, loss: 0.04637916758656502
step: 70, loss: 0.19251108169555664
step: 80, loss: 0.1201954334974289
step: 90, loss: 0.01074675191193819
step: 100, loss: 0.1001686081290245
step: 110, loss: 0.22538946568965912
step: 120, loss: 0.011440770700573921
step: 130, loss: 0.18573592603206635
step: 140, loss: 0.19450287520885468
step: 150, loss: 0.07801184803247452
step: 160, loss: 0.08352293819189072
step: 170, loss: 0.02071932889521122
step: 180, loss: 0.20794425904750824
step: 190, loss: 0.059463199228048325
step: 200, loss: 0.1473371833562851
step: 210, loss: 0.04618663340806961
/usr/local/lib/python3.7/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=0.6942148760330579, f1=0.7236580516898609,
best_f1=0.7065026362038664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09706173837184906
step: 10, loss: 0.2581113874912262
step: 20, loss: 0.030146002769470215
step: 30, loss: 0.10975385457277298
step: 40, loss: 0.07837317883968353
step: 50, loss: 0.04722217097878456
step: 60, loss: 0.09736263751983643
step: 70, loss: 0.08187068998813629
step: 80, loss: 0.026638120412826538
step: 90, loss: 0.0955851674079895
step: 100, loss: 0.05326550081372261
step: 110, loss: 0.019592449069023132
step: 120, loss: 0.033250708132982254
step: 130, loss: 0.14861832559108734
step: 140, loss: 0.03944804146885872
```

```
step: 150, loss: 0.12194876372814178
step: 160, loss: 0.03621995449066162
step: 170, loss: 0.05054721608757973
step: 180, loss: 0.17190445959568024
step: 190, loss: 0.04983041062951088
step: 200, loss: 0.09547209739685059
step: 210, loss: 0.006375199183821678
/usr/local/lib/python3.7/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=0.6836935166994106, f1=0.7192307692307692,
best_f1=0.7065026362038664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11779964715242386
step: 10, loss: 0.030814453959465027
step: 20, loss: 0.00906235072761774
step: 30, loss: 0.143998920917511
step: 40, loss: 0.04261419549584389
step: 50, loss: 0.04179776832461357
step: 60, loss: 0.13826951384544373
step: 70, loss: 0.005325130186975002
step: 80, loss: 0.07861624658107758
step: 90, loss: 0.20405150949954987
step: 100, loss: 0.10056523978710175
step: 110, loss: 0.031946275383234024
step: 120, loss: 0.014649314805865288
step: 130, loss: 0.045151904225349426
step: 140, loss: 0.05033107474446297
step: 150, loss: 0.02279522269964218
step: 160, loss: 0.19655704498291016
step: 170, loss: 0.025576557964086533
step: 180, loss: 0.13873687386512756
step: 190, loss: 0.13338303565979004
step: 200, loss: 0.016263501718640327
step: 210, loss: 0.05459632724523544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7027027027027026, f1=0.732824427480916,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03106847032904625
step: 10, loss: 0.09199357777833939
step: 20, loss: 0.01559009961783886
step: 30, loss: 0.03641920164227486
step: 40, loss: 0.07007426023483276
step: 50, loss: 0.12529370188713074
step: 60, loss: 0.10137752443552017
step: 70, loss: 0.0072745527140796185
step: 80, loss: 0.03954299911856651
step: 90, loss: 0.028658980503678322
step: 100, loss: 0.02278689481317997
step: 110, loss: 0.09674372524023056
step: 120, loss: 0.189444437623024
step: 130, loss: 0.01299175713211298
step: 140, loss: 0.004568816162645817
step: 150, loss: 0.10078274458646774
step: 160, loss: 0.014337075874209404
step: 170, loss: 0.06683003902435303
step: 180, loss: 0.0024149189703166485
step: 190, loss: 0.02679714746773243
step: 200, loss: 0.12179586291313171
step: 210, loss: 0.0143897021189332
/usr/local/lib/python3.7/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=0.7002012072434607, f1=0.7023809523809524,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.034984294325113297
step: 10, loss: 0.02136850357055664
step: 20, loss: 0.034079160541296005
step: 30, loss: 0.027563882991671562
step: 40, loss: 0.24717244505882263
step: 50, loss: 0.05591461807489395
step: 60, loss: 0.1850467026233673
step: 70, loss: 0.008765362203121185
step: 80, loss: 0.042906999588012695
step: 90, loss: 0.05165378749370575
step: 100, loss: 0.0054763504303991795
step: 110, loss: 0.0038129978347569704
step: 120, loss: 0.0154536422342062
step: 130, loss: 0.17404747009277344
step: 140, loss: 0.005015825852751732
step: 150, loss: 0.0508197620511055
step: 160, loss: 0.054583217948675156
```

step: 170, loss: 0.018308119848370552
step: 180, loss: 0.06325461715459824
step: 190, loss: 0.037114232778549194
step: 200, loss: 0.0370439738035202
step: 210, loss: 0.1178530901670456
/usr/local/lib/python3.7/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=0.6994106090373281, f1=0.7107750472589792,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10261663049459457
step: 10, loss: 0.16504406929016113
step: 20, loss: 0.038200005888938904
step: 30, loss: 0.0503661185503006
step: 40, loss: 0.1594875156879425
step: 50, loss: 0.039304058998823166
step: 60, loss: 0.04149383306503296
step: 70, loss: 0.05160951241850853
step: 80, loss: 0.05521959811449051
step: 90, loss: 0.259454607963562
step: 100, loss: 0.0445813350379467
step: 110, loss: 0.010293093509972095
step: 120, loss: 0.0036871498450636864
step: 130, loss: 0.01289005484431982
step: 140, loss: 0.023631054908037186
step: 150, loss: 0.011438174173235893
step: 160, loss: 0.01054002158343792
step: 170, loss: 0.11611924320459366
step: 180, loss: 0.016530562192201614
step: 190, loss: 0.1304536908864975
step: 200, loss: 0.023717103525996208
step: 210, loss: 0.19182482361793518
/usr/local/lib/python3.7/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=0.6964980544747081, f1=0.7166979362101313,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03468213230371475
step: 10, loss: 0.05248303711414337
step: 20, loss: 0.059729330241680145
step: 30, loss: 0.041743263602256775
step: 40, loss: 0.03955518826842308

```
step: 50, loss: 0.003569116583094001
step: 60, loss: 0.0053425077348947525
step: 70, loss: 0.02476450428366661
step: 80, loss: 0.0025753008667379618
step: 90, loss: 0.05387061461806297
step: 100, loss: 0.0020364911761134863
step: 110, loss: 0.0070436266250908375
step: 120, loss: 0.04312397539615631
step: 130, loss: 0.016548773273825645
step: 140, loss: 0.01526509691029787
step: 150, loss: 0.016846228390932083
step: 160, loss: 0.09027884900569916
step: 170, loss: 0.12538833916187286
step: 180, loss: 0.0068955314345657825
step: 190, loss: 0.14349789917469025
step: 200, loss: 0.009863914921879768
step: 210, loss: 0.030176708474755287
/usr/local/lib/python3.7/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=0.6805293005671078, f1=0.7087198515769945,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021679997444152832
step: 10, loss: 0.023026596754789352
step: 20, loss: 0.0027531499950876832
step: 30, loss: 0.07255984097719193
step: 40, loss: 0.039034172892570496
step: 50, loss: 0.005469315219670534
step: 60, loss: 0.030314680188894272
step: 70, loss: 0.0038655863609164953
step: 80, loss: 0.2966873347759247
step: 90, loss: 0.04508938267827034
step: 100, loss: 0.005801564082503319
step: 110, loss: 0.01565847173333168
step: 120, loss: 0.002087177475914359
step: 130, loss: 0.038352835923433304
step: 140, loss: 0.009780174121260643
step: 150, loss: 0.06575068086385727
step: 160, loss: 0.0026100901886820793
step: 170, loss: 0.057918015867471695
step: 180, loss: 0.03677577152848244
step: 190, loss: 0.008009077981114388
step: 200, loss: 0.0006890498334541917
step: 210, loss: 0.03694158419966698
/usr/local/lib/python3.7/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=0.6872586872586872, f1=0.7087198515769945,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006329046445898712
step: 10, loss: 0.016417428851127625
step: 20, loss: 0.0043663340620696545
step: 30, loss: 0.0021190079860389233
step: 40, loss: 0.03708299994468689
step: 50, loss: 0.007190322503447533
step: 60, loss: 0.02445082738995552
step: 70, loss: 0.0049741268157958984
step: 80, loss: 0.013815239071846008
step: 90, loss: 0.012210888788104057
step: 100, loss: 0.002190954051911831
step: 110, loss: 0.0033727672416716814
step: 120, loss: 0.10837350785732269
step: 130, loss: 0.016195023432374
step: 140, loss: 0.008769488893449306
step: 150, loss: 0.042814843356609344
step: 160, loss: 0.014615626074373722
step: 170, loss: 0.012057234533131123
step: 180, loss: 0.005449960939586163
step: 190, loss: 0.09043586254119873
step: 200, loss: 0.06219480186700821
step: 210, loss: 0.0787261351943016
/usr/local/lib/python3.7/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=0.6803278688524591, f1=0.688715953307393,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011179707944393158
step: 10, loss: 0.001994914375245571
step: 20, loss: 0.014596840366721153
step: 30, loss: 0.020270461216568947
step: 40, loss: 0.0030443647410720587
step: 50, loss: 0.08406086266040802
step: 60, loss: 0.008364999666810036
step: 70, loss: 0.1802024096250534
step: 80, loss: 0.009015152230858803
step: 90, loss: 0.03634748235344887
step: 100, loss: 0.19101449847221375
```

```
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 110, loss: 0.0007506810943596065
step: 120, loss: 0.015345069579780102
step: 130, loss: 0.07373097538948059
step: 140, loss: 0.008663316257297993
step: 150, loss: 0.0033035187516361475
step: 160, loss: 0.03150566667318344
step: 170, loss: 0.03439515456557274
step: 180, loss: 0.05562254413962364
step: 190, loss: 0.002045096829533577
step: 200, loss: 0.011996869929134846
step: 210, loss: 0.006587703246623278
/usr/local/lib/python3.7/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=0.6865671641791046, f1=0.7020109689213894,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007136638276278973
step: 10, loss: 0.002793537452816963
step: 20, loss: 0.05005449801683426
step: 30, loss: 0.2984643578529358
step: 40, loss: 0.021683502942323685
step: 50, loss: 0.013234330341219902
step: 60, loss: 0.024692559614777565
step: 70, loss: 0.05667111277580261
step: 80, loss: 0.21442970633506775
step: 90, loss: 0.16968537867069244
step: 100, loss: 0.002315917517989874
step: 110, loss: 0.06505290418863297
step: 120, loss: 0.017950521782040596
step: 130, loss: 0.0015214355662465096
step: 140, loss: 0.00042099173879250884
step: 150, loss: 0.015075121074914932
step: 160, loss: 0.03244593366980553
step: 170, loss: 0.01928180642426014
step: 180, loss: 0.007596186362206936
step: 190, loss: 0.01877141185104847
step: 200, loss: 0.0027037395630031824
step: 210, loss: 0.003469168208539486
/usr/local/lib/python3.7/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=0.6893787575150301, f1=0.6961538461538462,
best_f1=0.732824427480916
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0035364944487810135
step: 10, loss: 0.021036386489868164
step: 20, loss: 0.006391280330717564
step: 30, loss: 0.0053250547498464584
step: 40, loss: 0.0013207547599449754
step: 50, loss: 0.005892613437026739
step: 60, loss: 0.03341423347592354
step: 70, loss: 0.004011610988527536
step: 80, loss: 0.0014955730875954032
step: 90, loss: 0.0028148512355983257
step: 100, loss: 0.0018460279097780585
step: 110, loss: 0.021940693259239197
step: 120, loss: 0.06116955354809761
step: 130, loss: 0.031307633966207504
step: 140, loss: 0.006997418589890003
step: 150, loss: 0.0009038930875249207
step: 160, loss: 0.10976668447256088
step: 170, loss: 0.017998548224568367
step: 180, loss: 0.02192690409719944
step: 190, loss: 0.06671378761529922
step: 200, loss: 0.07588280737400055
step: 210, loss: 0.1342247724533081
/usr/local/lib/python3.7/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=0.6895238095238096, f1=0.708955223880597,
best_f1=0.732824427480916
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.")
2293it [00:02, 785.20it/s]
load_f1 = 0.6941838649155723
real_f1 = 0.6864564007421151
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 [00:09, 441.81it/s]
```

1.1.5 Walmart-Amazon - Running the matcher

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

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

      # Step 3: run the matcher
      !CUDA_VISIBLE_DEVICES=0 python matcher.py \
        --task Structured/Walmart-Amazon \
        --input_path input/input_small.jsonl \
```

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

Selected optimization level 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: 0.8722448945045471
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16849133372306824
step: 20, loss: 0.15284106135368347
step: 30, loss: 0.509779691696167
```

```
step: 40, loss: 0.26466697454452515
step: 50, loss: 0.309060662984848
step: 60, loss: 0.35999950766563416
step: 70, loss: 0.17114688456058502
step: 80, loss: 0.49708762764930725
step: 90, loss: 0.2267291098833084
step: 100, loss: 0.22146327793598175
step: 110, loss: 0.2387513518333435
step: 120, loss: 0.4149061143398285
step: 130, loss: 0.35510188341140747
step: 140, loss: 0.28650110960006714
step: 150, loss: 0.23548129200935364
step: 160, loss: 0.2006196528673172
step: 170, loss: 0.3406168222427368
step: 180, loss: 0.26045992970466614
step: 190, loss: 0.09383967518806458
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6787564766839378, f1=0.7286821705426356,
best_f1=0.7286821705426356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18938100337982178
step: 10, loss: 0.03324727714061737
step: 20, loss: 0.1764642298221588
step: 30, loss: 0.10502564907073975
step: 40, loss: 0.32930395007133484
step: 50, loss: 0.3172423243522644
step: 60, loss: 0.16377989947795868
step: 70, loss: 0.06789295375347137
step: 80, loss: 0.09700267016887665
step: 90, loss: 0.12496916204690933
step: 100, loss: 0.30894678831100464
step: 110, loss: 0.05060596019029617
step: 120, loss: 0.14285258948802948
step: 130, loss: 0.1409071832895279
step: 140, loss: 0.06107413023710251
step: 150, loss: 0.09374289214611053
step: 160, loss: 0.09281496703624725
step: 170, loss: 0.1211758404970169
step: 180, loss: 0.0867006853222847
step: 190, loss: 0.11714518815279007
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7653631284916201, f1=0.7875354107648725,
best_f1=0.7875354107648725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14861439168453217
step: 10, loss: 0.17689365148544312
step: 20, loss: 0.06473851203918457
step: 30, loss: 0.016749773174524307
step: 40, loss: 0.03525642305612564
step: 50, loss: 0.1636606752872467
step: 60, loss: 0.03460470214486122
step: 70, loss: 0.04602270573377609
step: 80, loss: 0.22249703109264374
step: 90, loss: 0.013766707852482796
step: 100, loss: 0.10519009083509445
step: 110, loss: 0.20415323972702026
step: 120, loss: 0.028281424194574356
step: 130, loss: 0.04956723004579544
step: 140, loss: 0.07077838480472565
step: 150, loss: 0.027578018605709076
step: 160, loss: 0.10467182099819183
step: 170, loss: 0.0779288187623024
step: 180, loss: 0.03365827351808548
step: 190, loss: 0.1956300437450409
/usr/local/lib/python3.7/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=0.7650000000000001, f1=0.7783251231527094,
best_f1=0.7875354107648725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09501676261425018
step: 10, loss: 0.03060663677752018
step: 20, loss: 0.049117907881736755
step: 30, loss: 0.03675013408064842
step: 40, loss: 0.008253036066889763
step: 50, loss: 0.0819634199142456
step: 60, loss: 0.018424000591039658
step: 70, loss: 0.027494732290506363
```

```
step: 80, loss: 0.05754426121711731
step: 90, loss: 0.04608244076371193
step: 100, loss: 0.013380749151110649
step: 110, loss: 0.013822178356349468
step: 120, loss: 0.05251634120941162
step: 130, loss: 0.244112029671669
step: 140, loss: 0.031179962679743767
step: 150, loss: 0.004668881185352802
step: 160, loss: 0.02404123544692993
step: 170, loss: 0.06208201125264168
step: 180, loss: 0.0819418877363205
step: 190, loss: 0.0143699049949646
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8042328042328043, f1=0.8054054054054054,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.057399120181798935
step: 10, loss: 0.007154824212193489
step: 20, loss: 0.054183971136808395
step: 30, loss: 0.012270680628716946
step: 40, loss: 0.13250137865543365
step: 50, loss: 0.014428353868424892
step: 60, loss: 0.010498270392417908
step: 70, loss: 0.0036923547741025686
step: 80, loss: 0.02183481492102146
step: 90, loss: 0.0044355206191539764
step: 100, loss: 0.05692274495959282
step: 110, loss: 0.0019943127408623695
step: 120, loss: 0.006691198330372572
step: 130, loss: 0.009161236695945263
step: 140, loss: 0.0028616103809326887
step: 150, loss: 0.03423531353473663
step: 160, loss: 0.07895714044570923
step: 170, loss: 0.030032843351364136
step: 180, loss: 0.07191938906908035
step: 190, loss: 0.3785865604877472
/usr/local/lib/python3.7/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=0.7989690721649483, f1=0.8, best_f1=0.8054054054054054
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08545234799385071
step: 10, loss: 0.004043381195515394
step: 20, loss: 0.005281398538500071
step: 30, loss: 0.0011509452015161514
step: 40, loss: 0.1273496150970459
step: 50, loss: 0.09008278697729111
step: 60, loss: 0.0036086784675717354
step: 70, loss: 0.024198273196816444
step: 80, loss: 0.06522060185670853
step: 90, loss: 0.014464118517935276
step: 100, loss: 0.0034947905223816633
step: 110, loss: 0.007759119849652052
step: 120, loss: 0.005134959705173969
step: 130, loss: 0.005467367824167013
step: 140, loss: 0.0027371644973754883
step: 150, loss: 0.0336458720266819
step: 160, loss: 0.01737060956656933
step: 170, loss: 0.03072960115969181
step: 180, loss: 0.018098700791597366
step: 190, loss: 0.048590365797281265
/usr/local/lib/python3.7/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=0.7645429362880887, f1=0.7830985915492957,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006584562361240387
step: 10, loss: 0.007788697257637978
step: 20, loss: 0.0014083960559219122
step: 30, loss: 0.09066595882177353
step: 40, loss: 0.12403390556573868
step: 50, loss: 0.23375919461250305
step: 60, loss: 0.006438985830545425
step: 70, loss: 0.0011860388331115246
step: 80, loss: 0.00313885067589581
step: 90, loss: 0.006598725914955139
step: 100, loss: 0.002209288300946355
step: 110, loss: 0.007324876729398966
step: 120, loss: 0.004456677008420229
step: 130, loss: 0.011854587122797966
step: 140, loss: 0.0008838780922815204
step: 150, loss: 0.02400788851082325
step: 160, loss: 0.0011650858214125037
```

```
step: 170, loss: 0.12077444046735764
step: 180, loss: 0.0008680368191562593
step: 190, loss: 0.014928828924894333
/usr/local/lib/python3.7/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=0.7665847665847665, f1=0.7969543147208122,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002347616944462061
step: 10, loss: 0.0045494744554162025
step: 20, loss: 0.06578334420919418
step: 30, loss: 0.009317459538578987
step: 40, loss: 0.10560019314289093
step: 50, loss: 0.014687635004520416
step: 60, loss: 0.002055011224001646
step: 70, loss: 0.0020275330170989037
step: 80, loss: 0.017450030893087387
step: 90, loss: 0.0016137135680764914
step: 100, loss: 0.024318579584360123
step: 110, loss: 0.017342137172818184
step: 120, loss: 0.0334247462451458
step: 130, loss: 0.0017480436945334077
step: 140, loss: 0.0013216926017776132
step: 150, loss: 0.007398709189146757
step: 160, loss: 0.0015565863577648997
step: 170, loss: 0.03088342398405075
step: 180, loss: 0.004767109639942646
step: 190, loss: 0.016905592754483223
/usr/local/lib/python3.7/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=0.7866323907455013, f1=0.8031088082901554,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012022528098896146
step: 10, loss: 0.04052618891000748
step: 20, loss: 0.01224115677177906
step: 30, loss: 0.0014739936450496316
step: 40, loss: 0.004305132664740086
step: 50, loss: 0.017013460397720337
step: 60, loss: 0.001174412784166634
step: 70, loss: 0.0015049547655507922
step: 80, loss: 0.0060089826583862305
```

```
step: 90, loss: 0.026718858629465103
step: 100, loss: 0.01126002985984087
step: 110, loss: 0.031682465225458145
step: 120, loss: 0.001303879776969552
step: 130, loss: 0.0031775429379194975
step: 140, loss: 0.0018697486957535148
step: 150, loss: 0.00454319640994072
step: 160, loss: 0.0027399882674217224
step: 170, loss: 0.028359852731227875
step: 180, loss: 0.04981926828622818
step: 190, loss: 0.0008258001180365682
/usr/local/lib/python3.7/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=0.7803617571059432, f1=0.8031496062992125,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001591630862094462
step: 10, loss: 0.018908681347966194
step: 20, loss: 0.0006346440059132874
step: 30, loss: 0.0339801162481308
step: 40, loss: 0.0004109203291591257
step: 50, loss: 0.0004978210781700909
step: 60, loss: 0.0020493550691753626
step: 70, loss: 0.009085682220757008
step: 80, loss: 0.010833886452019215
step: 90, loss: 0.0010108015267178416
step: 100, loss: 0.0018665492534637451
step: 110, loss: 0.06529682874679565
step: 120, loss: 0.011235237121582031
step: 130, loss: 0.001262524165213108
step: 140, loss: 0.0015482124872505665
step: 150, loss: 0.000432108121458441
step: 160, loss: 0.001267593470402062
step: 170, loss: 0.0033070959616452456
step: 180, loss: 0.00391465425491333
step: 190, loss: 0.0006330635515041649
/usr/local/lib/python3.7/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=0.7688564476885644, f1=0.7804878048780488,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008128737681545317
```

```
step: 10, loss: 0.008104688487946987
step: 20, loss: 0.05933830514550209
step: 30, loss: 0.050429921597242355
step: 40, loss: 0.0007310016662813723
step: 50, loss: 0.009320942685008049
step: 60, loss: 0.001946658012457192
step: 70, loss: 0.020344683900475502
step: 80, loss: 0.0003860041906591505
step: 90, loss: 0.10087762027978897
step: 100, loss: 0.0009348590392619371
step: 110, loss: 0.0009148152312263846
step: 120, loss: 0.0020639025606215
step: 130, loss: 0.001579072093591094
step: 140, loss: 0.00030279386555776
step: 150, loss: 0.0010527160484343767
step: 160, loss: 0.0005991411162540317
step: 170, loss: 0.0007306011393666267
step: 180, loss: 0.05747421085834503
step: 190, loss: 0.0020678481087088585
/usr/local/lib/python3.7/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=0.774025974025974, f1=0.7948051948051948,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00031881514587439597
step: 10, loss: 0.004892709665000439
step: 20, loss: 0.008074942044913769
step: 30, loss: 0.004145669285207987
step: 40, loss: 0.001968034775927663
step: 50, loss: 0.0020947030279785395
step: 60, loss: 0.002388907829299569
step: 70, loss: 0.0005070942570455372
step: 80, loss: 0.0014820171054452658
step: 90, loss: 0.004332855809479952
step: 100, loss: 0.0006563396309502423
step: 110, loss: 0.00033916381653398275
step: 120, loss: 0.12117507308721542
step: 130, loss: 0.0009010264766402543
step: 140, loss: 0.002762547926977277
step: 150, loss: 0.02226887084543705
step: 160, loss: 0.12761981785297394
step: 170, loss: 0.02216283045709133
step: 180, loss: 0.01314895786345005
step: 190, loss: 0.004583819303661585
/usr/local/lib/python3.7/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=0.781725888324873, f1=0.8120300751879699,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11899930238723755
step: 10, loss: 0.1744808405637741
step: 20, loss: 0.0011291527189314365
step: 30, loss: 0.00559830479323864
step: 40, loss: 0.0011620642617344856
step: 50, loss: 0.0070754545740783215
step: 60, loss: 0.0013107394333928823
step: 70, loss: 0.0016682022251188755
step: 80, loss: 0.0015312551986426115
step: 90, loss: 0.049845512956380844
step: 100, loss: 0.0015260810032486916
step: 110, loss: 0.08218157291412354
step: 120, loss: 0.0030421405099332333
step: 130, loss: 0.0023362054489552975
step: 140, loss: 0.002156514674425125
step: 150, loss: 0.0022778811398893595
step: 160, loss: 0.0059303343296051025
step: 170, loss: 0.000434401270467788
step: 180, loss: 0.0031659381929785013
step: 190, loss: 0.001713122008368373
/usr/local/lib/python3.7/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=0.7720207253886011, f1=0.7916666666666666,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03113461285829544
step: 10, loss: 0.0006327501614578068
step: 20, loss: 0.0011332231806591153
step: 30, loss: 0.012053866870701313
step: 40, loss: 0.057700756937265396
step: 50, loss: 0.0023563418071717024
step: 60, loss: 0.018268853425979614
step: 70, loss: 0.0018221080536022782
step: 80, loss: 0.0005095244268886745
step: 90, loss: 0.007585698273032904
step: 100, loss: 0.0013608250301331282
step: 110, loss: 0.0015665501123294234
step: 120, loss: 0.0016311903018504381
```

```
step: 130, loss: 0.0013283052248880267
step: 140, loss: 0.000330420327372849
step: 150, loss: 0.001078705070540309
step: 160, loss: 0.0007449883269146085
step: 170, loss: 0.0022344517055898905
step: 180, loss: 0.0012177417520433664
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.00026539029204286635
/usr/local/lib/python3.7/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=0.7745358090185677, f1=0.7913279132791328,
best_f1=0.8054054054054054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005649153143167496
step: 10, loss: 0.0004626455483958125
step: 20, loss: 0.0013506566174328327
step: 30, loss: 0.002601074753329158
step: 40, loss: 0.004178264643996954
step: 50, loss: 0.002760723466053605
step: 60, loss: 0.0010389169910922647
step: 70, loss: 0.0009650955908000469
step: 80, loss: 0.002131109358742833
step: 90, loss: 0.0017966488376259804
step: 100, loss: 0.0005319647607393563
step: 110, loss: 0.0004051878349855542
step: 120, loss: 0.0013020943151786923
step: 130, loss: 0.0005403292016126215
step: 140, loss: 0.07681401073932648
step: 150, loss: 0.0019212188199162483
step: 160, loss: 0.0056258924305438995
step: 170, loss: 0.0014569988707080483
step: 180, loss: 0.0014636347768828273
step: 190, loss: 0.04392508789896965
/usr/local/lib/python3.7/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=0.772972972972973, f1=0.7956403269754768,
best_f1=0.8054054054054054
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 [00:05, 376.66it/s]
load_f1 = 0.665
real_f1 = 0.6323185011709601
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 414.13it/s]
```

1.2 DA TEXTUAL

1.2.1 Abt-Buy - Running the matcher

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

```
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap  \
  --da span_shuffle \
  --summarize

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

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

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

Defaults for this optimization level are:
enabled   : True
opt_level   : 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: 0.8546686768531799
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22972211241722107
step: 20, loss: 0.15182551741600037
step: 30, loss: 0.23690347373485565
step: 40, loss: 0.3171442449092865
step: 50, loss: 0.3772675693035126
step: 60, loss: 0.4392097294330597
step: 70, loss: 0.3111962378025055
step: 80, loss: 0.24941357970237732
step: 90, loss: 0.3956091105937958
step: 100, loss: 0.22908757627010345
step: 110, loss: 0.15616203844547272
step: 120, loss: 0.4640842080116272
step: 130, loss: 0.3371104598045349
step: 140, loss: 0.2930046319961548
step: 150, loss: 0.09849527478218079
step: 160, loss: 0.18604615330696106
step: 170, loss: 0.1506420075893402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7258883248730964, f1=0.768472906403941,
best_f1=0.768472906403941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3502241373062134
step: 10, loss: 0.05691999942064285
step: 20, loss: 0.3071393668651581
step: 30, loss: 0.16528455913066864
step: 40, loss: 0.12590229511260986
step: 50, loss: 0.16811200976371765
step: 60, loss: 0.030340244993567467
step: 70, loss: 0.19523902237415314
```

```
step: 80, loss: 0.1293201446533203
step: 90, loss: 0.1805695742368698
step: 100, loss: 0.1151496022939682
step: 110, loss: 0.12168899923563004
step: 120, loss: 0.05277081951498985
step: 130, loss: 0.07059155404567719
step: 140, loss: 0.13805855810642242
step: 150, loss: 0.11886586993932724
step: 160, loss: 0.16196635365486145
step: 170, loss: 0.18732532858848572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7811764705882352, f1=0.7640449438202246,
best_f1=0.7640449438202246
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.036268942058086395
step: 10, loss: 0.058951858431100845
step: 20, loss: 0.08582853525876999
step: 30, loss: 0.14059381186962128
step: 40, loss: 0.006328528746962547
step: 50, loss: 0.1593443751335144
step: 60, loss: 0.17366738617420197
step: 70, loss: 0.043251048773527145
step: 80, loss: 0.06352889537811279
step: 90, loss: 0.12027067691087723
step: 100, loss: 0.03463292494416237
step: 110, loss: 0.012770034372806549
step: 120, loss: 0.012654440477490425
step: 130, loss: 0.08372952044010162
step: 140, loss: 0.0021057226695120335
step: 150, loss: 0.06319557875394821
step: 160, loss: 0.03066434897482395
step: 170, loss: 0.0850876122713089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8096385542168675, f1=0.7972350230414745,
```

```
best_f1=0.7972350230414745
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.041644878685474396
step: 10, loss: 0.19539354741573334
step: 20, loss: 0.0267979484051466
step: 30, loss: 0.04692286625504494
step: 40, loss: 0.01127129327505827
step: 50, loss: 0.016847314313054085
step: 60, loss: 0.09624911844730377
step: 70, loss: 0.009738314896821976
step: 80, loss: 0.02353746071457863
step: 90, loss: 0.014335744082927704
step: 100, loss: 0.01229174342006445
step: 110, loss: 0.013699590228497982
step: 120, loss: 0.19429774582386017
step: 130, loss: 0.021322568878531456
step: 140, loss: 0.010950678028166294
step: 150, loss: 0.0008983678417280316
step: 160, loss: 0.08822230994701385
step: 170, loss: 0.02053106017410755
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.821256038647343, f1=0.7999999999999999,
best_f1=0.7999999999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06951955705881119
step: 10, loss: 0.025492215529084206
step: 20, loss: 0.009999465197324753
step: 30, loss: 0.06493616849184036
step: 40, loss: 0.008986066095530987
step: 50, loss: 0.021316582337021828
step: 60, loss: 0.0011294392170384526
step: 70, loss: 0.01993756741285324
step: 80, loss: 0.005323351826518774
step: 90, loss: 0.005136576481163502
step: 100, loss: 0.029471831396222115
step: 110, loss: 0.10959041118621826
step: 120, loss: 0.060608528554439545
step: 130, loss: 0.004218824207782745
```

step: 140, loss: 0.058465324342250824
step: 150, loss: 0.0752285048365593
step: 160, loss: 0.004799015820026398
step: 170, loss: 0.0341406874358654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8292682926829268, f1=0.8262910798122066,
best_f1=0.8262910798122066
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04722658917307854
step: 10, loss: 0.09794881939888
step: 20, loss: 0.06409317255020142
step: 30, loss: 0.1005890890955925
step: 40, loss: 0.01503507886081934
step: 50, loss: 0.043089114129543304
step: 60, loss: 0.01477170642465353
step: 70, loss: 0.09863336384296417
step: 80, loss: 0.07656999677419662
step: 90, loss: 0.023449087515473366
step: 100, loss: 0.007813815958797932
step: 110, loss: 0.13118086755275726
step: 120, loss: 0.12894293665885925
step: 130, loss: 0.263436883687973
step: 140, loss: 0.007833714596927166
step: 150, loss: 0.09026152640581131
step: 160, loss: 0.0090136490762233373
step: 170, loss: 0.0016269948100671172
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8333333333333333, f1=0.821852731591449,
best_f1=0.821852731591449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023456912022083998
step: 10, loss: 0.0012851913925260305

```
step: 20, loss: 0.011072794906795025
step: 30, loss: 0.006993513088673353
step: 40, loss: 0.04440653696656227
step: 50, loss: 0.0037188460119068623
step: 60, loss: 0.08460398763418198
step: 70, loss: 0.00622683996334672
step: 80, loss: 0.031093375757336617
step: 90, loss: 0.07536628842353821
step: 100, loss: 0.09407613426446915
step: 110, loss: 0.002124917460605502
step: 120, loss: 0.22014650702476501
step: 130, loss: 0.16043469309806824
step: 140, loss: 0.05313839390873909
step: 150, loss: 0.031502071768045425
step: 160, loss: 0.007213693577796221
step: 170, loss: 0.11532099545001984
/usr/local/lib/python3.7/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=0.8325123152709358, f1=0.8028169014084506,
best_f1=0.821852731591449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022318333387374878
step: 10, loss: 0.01026833150535822
step: 20, loss: 0.005456042010337114
step: 30, loss: 0.050411880016326904
step: 40, loss: 0.017727313563227654
step: 50, loss: 0.002055979799479246
step: 60, loss: 0.004488872364163399
step: 70, loss: 0.018750421702861786
step: 80, loss: 0.0013002732302993536
step: 90, loss: 0.021561775356531143
step: 100, loss: 0.010738253593444824
step: 110, loss: 0.1321789175271988
step: 120, loss: 0.004889922682195902
step: 130, loss: 0.001990972552448511
step: 140, loss: 0.0021668691188097
step: 150, loss: 0.029653243720531464
step: 160, loss: 0.009126151911914349
step: 170, loss: 0.003825443796813488
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8366336633663366, f1=0.8184019370460048,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11134607344865799
step: 10, loss: 0.002812967635691166
step: 20, loss: 0.0015476912958547473
step: 30, loss: 0.016122179105877876
step: 40, loss: 0.03955591842532158
step: 50, loss: 0.008905555121600628
step: 60, loss: 0.0011537430109456182
step: 70, loss: 0.0336759090423584
step: 80, loss: 0.058569345623254776
step: 90, loss: 0.009883243590593338
step: 100, loss: 0.020833460614085197
step: 110, loss: 0.08750089257955551
step: 120, loss: 0.01291998103260994
step: 130, loss: 0.001167885959148407
step: 140, loss: 0.03635525703430176
step: 150, loss: 0.0034152877051383257
step: 160, loss: 0.00787558313459158
step: 170, loss: 0.00670506851747632
/usr/local/lib/python3.7/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=0.8298969072164948, f1=0.824742268041237,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008087717578746378
step: 10, loss: 0.04936622083187103
step: 20, loss: 0.0007004800136201084
step: 30, loss: 0.00020047435828018934
step: 40, loss: 0.036861080676317215
step: 50, loss: 0.03771153837442398
step: 60, loss: 0.0016167177818715572
step: 70, loss: 0.012378931045532227
step: 80, loss: 0.004432236775755882
step: 90, loss: 0.04782149940729141
step: 100, loss: 0.0018981820903718472
step: 110, loss: 0.007558753713965416
step: 120, loss: 0.016506172716617584
step: 130, loss: 0.00893388967961073
step: 140, loss: 0.010606624186038971
step: 150, loss: 0.0006659263162873685
```

```
step: 160, loss: 0.0026440212968736887
step: 170, loss: 0.005535438656806946
/usr/local/lib/python3.7/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=0.8325123152709358, f1=0.8254716981132074,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007626010337844491
step: 10, loss: 0.01369623001664877
step: 20, loss: 0.0012764474377036095
step: 30, loss: 0.0030296235345304012
step: 40, loss: 0.004546017851680517
step: 50, loss: 0.002388968365266919
step: 60, loss: 0.013827948831021786
step: 70, loss: 0.012850477360188961
step: 80, loss: 0.0010225152364000678
step: 90, loss: 0.005868355743587017
step: 100, loss: 0.000637431163340807
step: 110, loss: 0.003888980019837618
step: 120, loss: 0.002676402684301138
step: 130, loss: 0.007733955513685942
step: 140, loss: 0.08183201402425766
step: 150, loss: 0.001714809681288898
step: 160, loss: 0.0016179196536540985
step: 170, loss: 0.07241742312908173
/usr/local/lib/python3.7/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=0.8333333333333333, f1=0.796420581655481,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01447866391390562
step: 10, loss: 0.013557808473706245
step: 20, loss: 0.045421529561281204
step: 30, loss: 0.0029885785188525915
step: 40, loss: 0.00021440014825202525
step: 50, loss: 0.00017122318968176842
step: 60, loss: 0.001667490811087191
step: 70, loss: 0.009838646277785301
step: 80, loss: 0.05476515740156174
step: 90, loss: 0.0005126891774125397
step: 100, loss: 0.006218947935849428
step: 110, loss: 0.0007503421511501074
```

```
step: 120, loss: 0.05577843263745308
step: 130, loss: 0.03836905583739281
step: 140, loss: 0.006319731008261442
step: 150, loss: 0.01363910548388958
step: 160, loss: 0.0064566186629235744
step: 170, loss: 0.048927951604127884
/usr/local/lib/python3.7/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=0.8206388206388207, f1=0.8037383177570093,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042774114990606904
step: 10, loss: 0.10691113770008087
step: 20, loss: 0.02917102910578251
step: 30, loss: 0.0002763758529908955
step: 40, loss: 0.0001550452143419534
step: 50, loss: 0.01955699734389782
step: 60, loss: 0.000728771323338151
step: 70, loss: 0.0022126995027065277
step: 80, loss: 0.006251491140574217
step: 90, loss: 0.0027437913231551647
step: 100, loss: 0.00014642275345977396
step: 110, loss: 0.03961494192481041
step: 120, loss: 0.0008336291648447514
step: 130, loss: 0.012464862316846848
step: 140, loss: 0.00017917845980264246
step: 150, loss: 0.0008692040573805571
step: 160, loss: 0.0019082061480730772
step: 170, loss: 0.00856086052954197
/usr/local/lib/python3.7/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=0.8253164556962025, f1=0.8128078817733989,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002464211720507592
step: 10, loss: 0.00020630407379940152
step: 20, loss: 0.03175260126590729
step: 30, loss: 0.0004496254550758749
step: 40, loss: 0.001771307084709406
step: 50, loss: 0.001819383236579597
step: 60, loss: 0.00015507735952269286
step: 70, loss: 0.001371919410303235
```

```
step: 80, loss: 0.005662843585014343
step: 90, loss: 0.0008759048650972545
step: 100, loss: 0.0002746112586464733
step: 110, loss: 0.0007642555283382535
step: 120, loss: 0.0009974379790946841
step: 130, loss: 0.0031472390983253717
step: 140, loss: 0.0002856788341887295
step: 150, loss: 0.0002638689475134015
step: 160, loss: 0.024524148553609848
step: 170, loss: 0.015380761586129665
```
/usr/local/lib/python3.7/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=0.8270676691729323, f1=0.8164251207729469,
best_f1=0.8184019370460048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```
step: 0, loss: 0.0010071262950077653
step: 10, loss: 0.00787040963768959
step: 20, loss: 0.0051195137202739716
step: 30, loss: 0.0034758727997541428
step: 40, loss: 0.0024196510203182697
step: 50, loss: 0.0026476681232452393
step: 60, loss: 6.38441852061078e-05
step: 70, loss: 0.00023683221661485732
step: 80, loss: 0.00037156278267502785
step: 90, loss: 0.00303941173478961
step: 100, loss: 0.0015324825653806329
step: 110, loss: 0.0008665872737765312
step: 120, loss: 0.0031561313662678003
step: 130, loss: 0.003064091084524989
step: 140, loss: 0.044875480234622955
step: 150, loss: 0.045316871255636215
step: 160, loss: 0.00012899174180347472
step: 170, loss: 0.003608863800764084
```
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.")
epoch 15: dev_f1=0.8316831683168316, f1=0.8075117370892019,
best_f1=0.8184019370460048
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 [00:04, 465.05it/s]
load_f1 = 0.5368620037807184
real_f1 = 0.3374083129584352
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 427.77it/s]
```

1.3 DA DIRTY

1.3.1 DBLP-ACM - Running the matcher

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

```
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --da entry_swap  \
    --da span_shuffle \
    --summarize

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

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

```
Downloading: 100% 442/442 [00:00<00:00, 405kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 1.22MB/s]
Downloading: 100% 268M/268M [00:11<00:00, 23.2MB/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: 0.8222434520721436
step: 10, loss: 0.46680307388305664
step: 20, loss: 0.5538314580917358
step: 30, loss: 0.28843265771865845
step: 40, loss: 0.1441909670829773
step: 50, loss: 0.08887969702482224
step: 60, loss: 0.09471320360898972
step: 70, loss: 0.10106036812067032
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 80, loss: 0.20716264843940735
step: 90, loss: 0.021937349811196327
step: 100, loss: 0.12047069519758224
step: 110, loss: 0.14970660209655762
step: 120, loss: 0.03243102878332138
step: 130, loss: 0.01080082356929779
step: 140, loss: 0.005842032376676798
step: 150, loss: 0.2069014608860016
step: 160, loss: 0.11913081258535385
step: 170, loss: 0.10570122301578522
step: 180, loss: 0.008071869611740112
step: 190, loss: 0.02220604568719864
step: 200, loss: 0.010435281321406364
step: 210, loss: 0.05928003415465355
step: 220, loss: 0.00679635489359498
step: 230, loss: 0.015069318003952503
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9664429530201343, f1=0.9608938547486034,
best_f1=0.9608938547486034
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011029769666492939
step: 10, loss: 0.0018364939605817199
step: 20, loss: 0.002637198194861412
step: 30, loss: 0.0028905184008181095
step: 40, loss: 0.054110005497932434
step: 50, loss: 0.006927700247615576
step: 60, loss: 0.007363653741776943
step: 70, loss: 0.13467946648597717
step: 80, loss: 0.007958265021443367
step: 90, loss: 0.0054235332645475864
step: 100, loss: 0.017337601631879807
step: 110, loss: 0.17986717820167542
step: 120, loss: 0.0011142162838950753
step: 130, loss: 0.008651054464280605
step: 140, loss: 0.37822026014328003
step: 150, loss: 0.03712692856788635
step: 160, loss: 0.021265331655740738
step: 170, loss: 0.02851824276149273
step: 180, loss: 0.006711817812174559
step: 190, loss: 0.1254309117794037
step: 200, loss: 0.0041081286035478115
step: 210, loss: 0.030630722641944885
step: 220, loss: 0.0011825356632471085
step: 230, loss: 0.01240774616599083
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9751131221719457, f1=0.9651293588301463,
best_f1=0.9651293588301463
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03347625583410263
step: 10, loss: 0.016604691743850708
step: 20, loss: 0.010826285928487778
step: 30, loss: 0.008395510725677013
step: 40, loss: 0.01634388417005539
step: 50, loss: 0.044656869024038315
step: 60, loss: 0.002535826526582241
step: 70, loss: 0.0044591790065169334
step: 80, loss: 0.0060937777161598206
```

```
step: 90, loss: 0.0016745948232710361
step: 100, loss: 0.00149561429861933
step: 110, loss: 0.01337427832186222
step: 120, loss: 0.011026796884834766
step: 130, loss: 0.0030878065153956413
step: 140, loss: 0.0029306088108569384
step: 150, loss: 0.0016439581522718072
step: 160, loss: 0.004731596447527409
step: 170, loss: 0.01497694756835699
step: 180, loss: 0.018210140988230705
step: 190, loss: 0.00471072643995285
step: 200, loss: 0.01586635783314705
step: 210, loss: 0.0037078510504215956
step: 220, loss: 0.006726751569658518
step: 230, loss: 0.09312189370393753
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9752252252252253, f1=0.9730337078651685,
best_f1=0.9730337078651685
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0073888846673071384
step: 10, loss: 0.003029066137969494
step: 20, loss: 0.0022056428715586662
step: 30, loss: 0.002403038553893566
step: 40, loss: 0.007462657056748867
step: 50, loss: 0.006625502370297909
step: 60, loss: 0.0010288608027622104
step: 70, loss: 0.031961824744939804
step: 80, loss: 0.08979557454586029
step: 90, loss: 0.002202109433710575
step: 100, loss: 0.0025761821307241917
step: 110, loss: 0.01315730158239603
step: 120, loss: 0.004533371888101101
step: 130, loss: 0.001814217888750136
step: 140, loss: 0.0002708998217713088
step: 150, loss: 0.0011290735565125942
step: 160, loss: 0.005156898405402899
step: 170, loss: 0.0014663658803328872
step: 180, loss: 0.05139859393239021
step: 190, loss: 0.004718919284641743
step: 200, loss: 0.07732166349887848
```

```
step: 210, loss: 0.07023531198501587
step: 220, loss: 0.0011824631365016103
step: 230, loss: 0.0033649897668510675
/usr/local/lib/python3.7/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=0.9732739420935412, f1=0.96875, best_f1=0.9730337078651685
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004874141421169043
step: 10, loss: 0.005936859641224146
step: 20, loss: 0.002075977623462677
step: 30, loss: 0.002338983351364732
step: 40, loss: 0.0013719944981858134
step: 50, loss: 0.001910269958898425
step: 60, loss: 0.0008162384619936347
step: 70, loss: 0.00033889617770910263
step: 80, loss: 0.001412736950442195
step: 90, loss: 0.028157396242022514
step: 100, loss: 0.0034717798698693514
step: 110, loss: 0.0020457888022065163
step: 120, loss: 0.08785632997751236
step: 130, loss: 0.08584283292293549
step: 140, loss: 0.0028843802865594625
step: 150, loss: 0.0006559204193763435
step: 160, loss: 0.0657804012298584
step: 170, loss: 0.024038583040237427
step: 180, loss: 0.0019545990508049726
step: 190, loss: 0.0005745073431171477
step: 200, loss: 0.002613178687170148
step: 210, loss: 0.0011195102706551552
step: 220, loss: 0.0008518461836501956
step: 230, loss: 0.0008631121017970145
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9787234042553192, f1=0.9743589743589743,
best_f1=0.9743589743589743
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0022382044699043036
step: 10, loss: 0.0003050228697247803
```

```
step: 20, loss: 0.0010860078036785126
step: 30, loss: 0.0009295390918850899
step: 40, loss: 0.0016819294542074203
step: 50, loss: 0.03403014317154884
step: 60, loss: 0.013023835606873035
step: 70, loss: 0.004755641333758831
step: 80, loss: 0.00646429555490613
step: 90, loss: 0.0010401987237855792
step: 100, loss: 0.0024832666385918856
step: 110, loss: 0.03424585610628128
step: 120, loss: 0.0005744716036133468
step: 130, loss: 0.0065298620611429214
step: 140, loss: 0.009977913461625576
step: 150, loss: 0.0010733188828453422
step: 160, loss: 0.05657525733113289
step: 170, loss: 0.0005795176839455962
step: 180, loss: 0.0004287216579541564
step: 190, loss: 0.008744806982576847
step: 200, loss: 0.01369442418217659
step: 210, loss: 0.002053579781204462
step: 220, loss: 0.00037536583840847015
step: 230, loss: 0.00035256799310445786
/usr/local/lib/python3.7/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=0.9775280898876404, f1=0.9753363228699552,
best_f1=0.9743589743589743
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10106216371059418
step: 10, loss: 0.00020624400349333882
step: 20, loss: 0.0004777238063979894
step: 30, loss: 0.002476408611983061
step: 40, loss: 0.008098593913018703
step: 50, loss: 0.0005430838209576905
step: 60, loss: 0.07325389981269836
step: 70, loss: 0.0005407187272794545
step: 80, loss: 0.0006009888020344079
step: 90, loss: 0.001997290411964059
step: 100, loss: 0.0003963318595197052
step: 110, loss: 0.002310530049726367
step: 120, loss: 0.0004781304451171309
step: 130, loss: 0.00047466211253777146
step: 140, loss: 0.00017890348681248724
step: 150, loss: 0.0025238092057406902
step: 160, loss: 0.0002324461383977905
step: 170, loss: 0.001304080942645669
```

```
step: 180, loss: 0.001990898745134473
step: 190, loss: 0.006653878837823868
step: 200, loss: 0.015247893519699574
step: 210, loss: 0.00038669182686135173
step: 220, loss: 0.01979469507932663
step: 230, loss: 0.020738031715154648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9753363228699552,
best_f1=0.9753363228699552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04157349839806557
step: 10, loss: 0.07227130979299545
step: 20, loss: 0.00023807563411537558
step: 30, loss: 0.0008436960051767528
step: 40, loss: 0.0005680356407538056
step: 50, loss: 0.001534111681394279
step: 60, loss: 0.00034489246900193393
step: 70, loss: 0.00020355470769573003
step: 80, loss: 0.0004784634511452168
step: 90, loss: 0.00013456594024319202
step: 100, loss: 0.00018258923955727369
step: 110, loss: 0.0004429776454344392
step: 120, loss: 0.00012544992205221206
step: 130, loss: 0.009243826381862164
step: 140, loss: 0.005296381656080484
step: 150, loss: 0.00011059060489060357
step: 160, loss: 0.00031157402554526925
step: 170, loss: 0.00023953733034431934
step: 180, loss: 0.00037741236155852675
step: 190, loss: 0.0007506947731599212
step: 200, loss: 0.010729274712502956
step: 210, loss: 0.00010931353608611971
step: 220, loss: 0.00019586387497838587
step: 230, loss: 0.04712804779410362
/usr/local/lib/python3.7/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=0.9798206278026906, f1=0.9753914988814317,
best_f1=0.9753363228699552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.469933527521789e-05
step: 10, loss: 0.0005348905106075108
step: 20, loss: 0.00021050703071523458
step: 30, loss: 0.0010599188972264528
step: 40, loss: 5.1399721996858716e-05
step: 50, loss: 8.826227713143453e-05
step: 60, loss: 9.686980047263205e-05
step: 70, loss: 0.0001146387803601101
step: 80, loss: 0.021348310634493828
step: 90, loss: 0.00087047566194408727
step: 100, loss: 0.0004133158072363585
step: 110, loss: 6.771599146304652e-05
step: 120, loss: 0.02363039180636406
step: 130, loss: 6.773616041755304e-05
step: 140, loss: 0.03429966792464256
step: 150, loss: 9.361132106278092e-05
step: 160, loss: 0.00012033965322189033
step: 170, loss: 0.00010591370664769784
step: 180, loss: 0.0018136819126084447
step: 190, loss: 0.0017469098092988133
step: 200, loss: 0.00025859783636406064
step: 210, loss: 0.00041551515460014343
step: 220, loss: 0.018732082098722458
step: 230, loss: 0.011578081175684929
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9810055865921787, f1=0.9755555555555556,
best_f1=0.9755555555555556
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013834076526110913
step: 10, loss: 0.00013286079047247767
step: 20, loss: 0.00030803203117102385
step: 30, loss: 0.00017241996829397976
step: 40, loss: 5.3803665650775656e-05
step: 50, loss: 0.0014837387716397643
step: 60, loss: 0.1454620659351349
step: 70, loss: 0.000566199014429003
step: 80, loss: 0.0011488223681226373
step: 90, loss: 0.006484219804406166

```
step: 100, loss: 0.0010519381612539291
step: 110, loss: 0.0001783082989277318
step: 120, loss: 0.02751731500029564
step: 130, loss: 0.0007780352607369423
step: 140, loss: 0.05708368495106697
step: 150, loss: 0.0001200955593958497
step: 160, loss: 0.00033914620871655643
step: 170, loss: 9.540648898109794e-05
step: 180, loss: 0.0003195136087015271
step: 190, loss: 0.00013288381160236895
step: 200, loss: 0.00015438697300851345
step: 210, loss: 5.6245287851197645e-05
step: 220, loss: 0.00023294027778320014
step: 230, loss: 0.0005431731115095317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9821029082774049, f1=0.9723145071982282,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.162796828197315e-05
step: 10, loss: 0.00010869171092053875
step: 20, loss: 3.6683562939288095e-05
step: 30, loss: 9.159486944554374e-05
step: 40, loss: 0.017334578558802605
step: 50, loss: 0.00012005286407656968
step: 60, loss: 7.113642641343176e-05
step: 70, loss: 0.0001100455192499794
step: 80, loss: 0.00012488551146816462
step: 90, loss: 5.1270428230054677e-05
step: 100, loss: 0.00019510489073581994
step: 110, loss: 0.04071544483304024
step: 120, loss: 0.0001835203729569912
step: 130, loss: 4.837643791688606e-05
step: 140, loss: 0.000254010665230453
step: 150, loss: 3.876311893691309e-05
step: 160, loss: 6.502831820398569e-05
step: 170, loss: 0.012849629856646061
step: 180, loss: 0.00011180694855283946
step: 190, loss: 9.52112750383094e-05
step: 200, loss: 7.715893298154697e-05
step: 210, loss: 5.3777759603690356e-05
```

```
step: 220, loss: 4.853786958847195e-05
step: 230, loss: 0.033059969544410706
/usr/local/lib/python3.7/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=0.9821029082774049, f1=0.9712389380530975,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012996501754969358
step: 10, loss: 0.003828381886705756
step: 20, loss: 3.669234865810722e-05
step: 30, loss: 0.0007922141812741756
step: 40, loss: 6.913063407409936e-05
step: 50, loss: 0.0009843669831752777
step: 60, loss: 0.0006396552198566496
step: 70, loss: 0.05593792721629143
step: 80, loss: 0.0003464682085905224
step: 90, loss: 0.00012338533997535706
step: 100, loss: 0.00010055655002361163
step: 110, loss: 0.00210681464523077
step: 120, loss: 0.0010068658739328384
step: 130, loss: 0.008426508866250515
step: 140, loss: 5.125333700561896e-05
step: 150, loss: 0.00039998433203436434
step: 160, loss: 0.024099543690681458
step: 170, loss: 7.858887693146244e-05
step: 180, loss: 0.00010866076627280563
step: 190, loss: 4.6171433496056125e-05
step: 200, loss: 0.006524808704853058
step: 210, loss: 6.306263821898028e-05
step: 220, loss: 0.026936829090118408
step: 230, loss: 0.0020904988050460815
/usr/local/lib/python3.7/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=0.9820224719101124, f1=0.972129319955407,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.368846541387029e-05
step: 10, loss: 5.6414184655295685e-05
step: 20, loss: 0.00013464085350278765
step: 30, loss: 0.0036962535232305527
step: 40, loss: 0.0004405368526931852
step: 50, loss: 0.00013746116019319743
```

```
step: 60, loss: 5.444213456939906e-05
step: 70, loss: 6.71740563120693e-05
step: 80, loss: 6.244658288778737e-05
step: 90, loss: 5.6879533076426014e-05
step: 100, loss: 0.0001457155158277601
step: 110, loss: 0.00012387809692882001
step: 120, loss: 6.648476119153202e-05
step: 130, loss: 0.003916235640645027
step: 140, loss: 0.01513738464564085
step: 150, loss: 0.00010507369006518275
step: 160, loss: 0.00012860474816989154
step: 170, loss: 8.016196079552174e-05
step: 180, loss: 6.726918218191713e-05
step: 190, loss: 5.791211879113689e-05
step: 200, loss: 0.0003291832108516246
step: 210, loss: 0.0005392152233980596
step: 220, loss: 5.305834565660916e-05
step: 230, loss: 4.113122122362256e-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=0.9796380090497738, f1=0.9764837625979844,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.265836418606341e-05
step: 10, loss: 2.289493932039477e-05
step: 20, loss: 0.003982621245086193
step: 30, loss: 6.244542601052672e-05
step: 40, loss: 2.527519791328814e-05
step: 50, loss: 8.899944077711552e-05
step: 60, loss: 5.424064875114709e-05
step: 70, loss: 9.000091085908934e-05
step: 80, loss: 9.203882655128837e-05
step: 90, loss: 0.0001352810359094292
step: 100, loss: 0.00696075614541769
step: 110, loss: 6.893358658999205e-05
step: 120, loss: 6.220197974471375e-05
step: 130, loss: 0.00026943700504489243
step: 140, loss: 7.661274139536545e-05
step: 150, loss: 5.947132740402594e-05
step: 160, loss: 3.637748886831105e-05
step: 170, loss: 3.3797896321630105e-05
step: 180, loss: 7.135045598261058e-05
step: 190, loss: 0.00020905601559206843
step: 200, loss: 3.0981882446212694e-05
step: 210, loss: 0.0001604818389751017
```

```
step: 220, loss: 7.503340020775795e-05
step: 230, loss: 1.9326455003465526e-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=0.9740112994350283,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.8635669625364244e-05
step: 10, loss: 0.00016130109725054353
step: 20, loss: 4.436869130586274e-05
step: 30, loss: 9.33894989429973e-05
step: 40, loss: 5.258223973214626e-05
step: 50, loss: 6.0704660427290946e-05
step: 60, loss: 4.099080615560524e-05
step: 70, loss: 4.105442712898366e-05
step: 80, loss: 0.01556189265102148
step: 90, loss: 3.778657628572546e-05
step: 100, loss: 6.251138256629929e-05
step: 110, loss: 0.00017589234630577266
step: 120, loss: 4.495331450016238e-05
step: 130, loss: 0.0006057508289813995
step: 140, loss: 0.0002540012646932155
step: 150, loss: 5.6005967053351924e-05
step: 160, loss: 4.1888022678904235e-05
step: 170, loss: 2.6631039872881956e-05
step: 180, loss: 6.532250699819997e-05
step: 190, loss: 0.00010498063784325495
step: 200, loss: 3.3473352232249454e-05
step: 210, loss: 0.0031294531654566526
step: 220, loss: 0.0043316255532358885
step: 230, loss: 4.843372880714014e-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=0.9784824462061155, f1=0.9764837625979844,
best_f1=0.9723145071982282
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
```

```
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:08, 298.84it/s]
load_f1 = 0.9798657718120806
real_f1 = 0.9754464285714286
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, 361.36it/s]
```

1.3.2 DBLP-GoogleScholar - Running the matcher

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

```
  --da attr_del \
  --summarize

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

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

Selected optimization level 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: 0.7954671382904053
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4445924162864685
step: 20, loss: 0.49446630477905273
step: 30, loss: 0.42956772446632385
step: 40, loss: 0.30998095870018005
step: 50, loss: 0.20082195103168488
step: 60, loss: 0.08853202313184738
step: 70, loss: 0.06397716701030731
step: 80, loss: 0.13763457536697388
step: 90, loss: 0.12564155459403992
step: 100, loss: 0.19194185733795166
step: 110, loss: 0.14046278595924377
step: 120, loss: 0.034680113196372986
step: 130, loss: 0.03185919672250748
step: 140, loss: 0.1936059445142746
step: 150, loss: 0.035813722759485245
step: 160, loss: 0.16118758916854858
step: 170, loss: 0.0642261877655983
step: 180, loss: 0.0955386608839035
step: 190, loss: 0.03448177129030228
step: 200, loss: 0.07728838920593262
step: 210, loss: 0.07306677848100662
step: 220, loss: 0.057238344103097916
step: 230, loss: 0.15005844831466675
step: 240, loss: 0.05475986748933792
step: 250, loss: 0.037070270627737045
step: 260, loss: 0.052611276507377625
step: 270, loss: 0.014636781066656113
step: 280, loss: 0.12790504097938538
step: 290, loss: 0.06596971303224564
step: 300, loss: 0.0976877212524414
step: 310, loss: 0.14779691398143768
step: 320, loss: 0.02099895291030407
step: 330, loss: 0.13652601838111877
step: 340, loss: 0.12085016071796417
step: 350, loss: 0.07242146879434586
step: 360, loss: 0.029464976862072945
step: 370, loss: 0.10592871904373169
step: 380, loss: 0.11515931785106659
step: 390, loss: 0.015357553958892822
step: 400, loss: 0.019252579659223557
step: 410, loss: 0.04301767796278

```
step: 420, loss: 0.019740913063287735
step: 430, loss: 0.05242788419127464
step: 440, loss: 0.1262032389640808
step: 450, loss: 0.049835868179798126
step: 460, loss: 0.1313575804233551
step: 470, loss: 0.13109175860881805
step: 480, loss: 0.3897755742073059
step: 490, loss: 0.018269410356879234
step: 500, loss: 0.007911540567874908
step: 510, loss: 0.08088318258523941
step: 520, loss: 0.09854359179735184
step: 530, loss: 0.15057356655597687
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9300797747536368, f1=0.9299719887955181,
best_f1=0.9299719887955181
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13415813446044922
step: 10, loss: 0.18745213747024536
step: 20, loss: 0.08263368904590607
step: 30, loss: 0.046848323196172714
step: 40, loss: 0.003066032426431775
step: 50, loss: 0.13716158270835876
step: 60, loss: 0.1531345248222351
step: 70, loss: 0.19621029496192932
step: 80, loss: 0.03955121710896492
step: 90, loss: 0.010185355320572853
step: 100, loss: 0.29560747742652893
step: 110, loss: 0.09206690639257431
step: 120, loss: 0.09105543792247772
step: 130, loss: 0.02460724115371704
step: 140, loss: 0.025857388973236084
step: 150, loss: 0.06401024013757706
step: 160, loss: 0.018523452803492546
step: 170, loss: 0.07589440047740936
step: 180, loss: 0.03987827152013779
step: 190, loss: 0.011265140026807785
step: 200, loss: 0.05959209427237511
step: 210, loss: 0.005313739646226168
step: 220, loss: 0.09047719836235046
step: 230, loss: 0.01496612373739481
```

```
step: 240, loss: 0.13804318010807037
step: 250, loss: 0.02121000736951828
step: 260, loss: 0.08822104334831238
step: 270, loss: 0.05781498923897743
step: 280, loss: 0.09356500953435898
step: 290, loss: 0.02466561459004879
step: 300, loss: 0.008588681928813457
step: 310, loss: 0.048127681016922
step: 320, loss: 0.19543275237083435
step: 330, loss: 0.03143876791000366
step: 340, loss: 0.0789499431848526
step: 350, loss: 0.06430266052484512
step: 360, loss: 0.047820866107940674
step: 370, loss: 0.014618253335356712
step: 380, loss: 0.07356013357639313
step: 390, loss: 0.03931891545653343
step: 400, loss: 0.0781404972076416
step: 410, loss: 0.0031705403234809637
step: 420, loss: 0.08445867896080017
step: 430, loss: 0.030302949249744415
step: 440, loss: 0.016285419464111328
step: 450, loss: 0.022046543657779694
step: 460, loss: 0.18955054879188538
step: 470, loss: 0.038543786853551865
step: 480, loss: 0.18909917771816254
step: 490, loss: 0.012596861459314823
step: 500, loss: 0.0347476489841938
step: 510, loss: 0.08462020009756088
step: 520, loss: 0.07563892751932144
step: 530, loss: 0.1312086284160614
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9453053783044667, f1=0.9344188150158298,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05038445070385933
step: 10, loss: 0.035032711923122406
step: 20, loss: 0.11585860699415207
step: 30, loss: 0.18293620645999908
step: 40, loss: 0.0036195283755660057
step: 50, loss: 0.010764527134597301
```

```
step: 60, loss: 0.028379814699292183
step: 70, loss: 0.0186567734926939
step: 80, loss: 0.0012030770303681493
step: 90, loss: 0.006299525965005159
step: 100, loss: 0.043019525706768036
step: 110, loss: 0.002151880646124482
step: 120, loss: 0.013414496555924416
step: 130, loss: 0.04280843585729599
step: 140, loss: 0.019613312555733967
step: 150, loss: 0.001890919404104352
step: 160, loss: 0.002970536006614566
step: 170, loss: 0.0056956857442855835
step: 180, loss: 0.06946120411157608
step: 190, loss: 0.05746178328990936
step: 200, loss: 0.004643015097826719
step: 210, loss: 0.14419907331466675
step: 220, loss: 0.014647362753748894
step: 230, loss: 0.017553754150867462
step: 240, loss: 0.011602314189076424
step: 250, loss: 0.015798168256878853
step: 260, loss: 0.004324715584516525
step: 270, loss: 0.0015455657849088311
step: 280, loss: 0.015269717201590538
step: 290, loss: 0.023750215768814087
step: 300, loss: 0.08553830534219742
step: 310, loss: 0.18630295991897583
step: 320, loss: 0.1265580803155899
step: 330, loss: 0.0017331927083432674
step: 340, loss: 0.008919540792703629
step: 350, loss: 0.0210714228451252
step: 360, loss: 0.0044242083095014095
step: 370, loss: 0.01228855550289154
step: 380, loss: 0.00638435548171401
step: 390, loss: 0.03928859531879425
step: 400, loss: 0.014181378297507763
step: 410, loss: 0.004293524194508791
step: 420, loss: 0.07103630155324936
step: 430, loss: 0.09171586483716965
step: 440, loss: 0.03998931124806404
step: 450, loss: 0.0773925930261612
step: 460, loss: 0.041544925421476364
step: 470, loss: 0.011213162913918495
step: 480, loss: 0.006314816419035196
step: 490, loss: 0.008867637254297733
step: 500, loss: 0.1112065464258194
step: 510, loss: 0.0015524604823440313
step: 520, loss: 0.0030213582795113325
step: 530, loss: 0.03292607143521309
```

```
/usr/local/lib/python3.7/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=0.9385113268608415, f1=0.9304911955514366,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0081351669950523853
step: 10, loss: 0.005012051668018103
step: 20, loss: 0.044995978474617004
step: 30, loss: 0.014544559642672539
step: 40, loss: 0.002427920000627637
step: 50, loss: 0.01435715239495039
step: 60, loss: 0.0008132434450089931
step: 70, loss: 0.0019484392832964659
step: 80, loss: 0.01040623988956213
step: 90, loss: 0.01981007680296898
step: 100, loss: 0.011172587983310223
step: 110, loss: 0.0012624879600480199
step: 120, loss: 0.0014054153580218554
step: 130, loss: 0.12115409225225449
step: 140, loss: 0.24797934293746948
step: 150, loss: 0.0053691864013671875
step: 160, loss: 0.005287630949169397
step: 170, loss: 0.0066828918643295765
step: 180, loss: 0.017648737877607346
step: 190, loss: 0.0074978661723434925
step: 200, loss: 0.0006729031447321177
step: 210, loss: 0.06533142924308777
step: 220, loss: 0.004240177571773529
step: 230, loss: 0.34186917543411255
step: 240, loss: 0.03659653663635254
step: 250, loss: 0.002258839551359415
step: 260, loss: 0.0992899015545845
step: 270, loss: 0.05738629773259163
step: 280, loss: 0.010546263307332993
step: 290, loss: 0.02402864582836628
step: 300, loss: 0.0015671112341806293
step: 310, loss: 0.0032866974361240864
step: 320, loss: 0.007781027816236019
step: 330, loss: 0.16950131952762604
step: 340, loss: 0.15569552779197693
step: 350, loss: 0.049906276166439056
step: 360, loss: 0.010163742117583752
step: 370, loss: 0.09466435760259628
step: 380, loss: 0.009646104648709297
step: 390, loss: 0.027657408267259598
```

```
step: 400, loss: 0.010276032611727715
step: 410, loss: 0.007325673010200262
step: 420, loss: 0.0003969725512433797
step: 430, loss: 0.005655454006046057
step: 440, loss: 0.050215281546115875
step: 450, loss: 0.02378562092781067
step: 460, loss: 0.0052670990116894245
step: 470, loss: 0.0038020354695618153
step: 480, loss: 0.0048904744908213615
step: 490, loss: 0.052950188517570496
step: 500, loss: 0.014711259864270687
step: 510, loss: 0.009431489743292332
step: 520, loss: 0.011294934898614883
step: 530, loss: 0.0009095969144254923
/usr/local/lib/python3.7/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=0.9408502772643254, f1=0.9314814814814815,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006689819507300854
step: 10, loss: 0.11182057857513428
step: 20, loss: 0.02763666957616806
step: 30, loss: 0.024284830316901207
step: 40, loss: 0.09106539934873581
step: 50, loss: 0.003970021847635508
step: 60, loss: 0.019215364009141922
step: 70, loss: 0.0015455594984814525
step: 80, loss: 0.0011792141012847424
step: 90, loss: 0.2890245318412781
step: 100, loss: 0.0009412094368599355
step: 110, loss: 0.003819490084424615
step: 120, loss: 0.008447183296084404
step: 130, loss: 0.010378282517194748
step: 140, loss: 0.002634556731209159
step: 150, loss: 0.0020135699305683374
step: 160, loss: 0.16783694922924042
step: 170, loss: 0.030165525153279305
step: 180, loss: 0.020017258822917938
step: 190, loss: 0.0016017199959605932
step: 200, loss: 0.10291419923305511
step: 210, loss: 0.000811011355835944
step: 220, loss: 0.0003973742132075131
step: 230, loss: 0.0035954706836491823
step: 240, loss: 0.005802084226161241  5
step: 250, loss: 0.0006804258446209133
```

```
step: 260, loss: 0.02766539715230465
step: 270, loss: 0.008999642916023731
step: 280, loss: 0.021035367622971535
step: 290, loss: 0.006982394494116306
step: 300, loss: 0.006646405905485153
step: 310, loss: 0.0005180027219466865
step: 320, loss: 0.001340094138868153
step: 330, loss: 0.0024219953920692205
step: 340, loss: 0.007250686641782522
step: 350, loss: 0.03569691255688667
step: 360, loss: 0.033982083201408386
step: 370, loss: 0.006624253466725349
step: 380, loss: 0.07830997556447983
step: 390, loss: 0.0035900536458939314
step: 400, loss: 0.0042751291766676273
step: 410, loss: 0.002364327432587743
step: 420, loss: 0.002452347194775939
step: 430, loss: 0.014445138163864613
step: 440, loss: 0.06214035302400589
step: 450, loss: 0.10369081795215607
step: 460, loss: 0.05900858715176582
step: 470, loss: 0.004195588640868664
step: 480, loss: 0.014535021036863327
step: 490, loss: 0.0029008807614445686
step: 500, loss: 0.013301133178174496
step: 510, loss: 0.005405465606600046
step: 520, loss: 0.004386219196021557
step: 530, loss: 0.013089796528220177
/usr/local/lib/python3.7/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=0.9437470943747095, f1=0.9388322520852641,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0045365579426288605
step: 10, loss: 0.002436721930280328
step: 20, loss: 0.055936917662620544
step: 30, loss: 0.0023655646946281195
step: 40, loss: 0.0007508864509873092
step: 50, loss: 0.001640256610698998
step: 60, loss: 0.0013189382152631879
step: 70, loss: 0.052186932414770126
step: 80, loss: 0.004528752062469721
step: 90, loss: 0.03031487576663494
step: 100, loss: 0.1689150482416153
step: 110, loss: 0.04710659384727478
```

```
step: 120, loss: 0.049121782183647156
step: 130, loss: 0.0042314669117331505
step: 140, loss: 0.006197468377649784
step: 150, loss: 0.0038186698220670223
step: 160, loss: 0.0016178227961063385
step: 170, loss: 0.0014090380864217877
step: 180, loss: 0.00010147284774575382
step: 190, loss: 0.0029461337253451347
step: 200, loss: 0.0004074170719832182
step: 210, loss: 0.02599596604704857
step: 220, loss: 0.0009315226343460381
step: 230, loss: 0.00014193644165061414
step: 240, loss: 0.0012232058215886354
step: 250, loss: 0.012488119304180145
step: 260, loss: 0.0007758757565170527
step: 270, loss: 0.00010394649871159345
step: 280, loss: 0.00019262680143583566
step: 290, loss: 0.00804778654128313
step: 300, loss: 0.0008417365606874228
step: 310, loss: 0.007160323206335306
step: 320, loss: 0.11373239755630493
step: 330, loss: 0.0012970705283805728
step: 340, loss: 0.020082345232367516
step: 350, loss: 0.08244217187166214
step: 360, loss: 0.006536643486469984
step: 370, loss: 0.012396254576742649
step: 380, loss: 0.0103090088814497
step: 390, loss: 0.004162504803389311
step: 400, loss: 0.00020018103532493114
step: 410, loss: 0.009284068830311298
step: 420, loss: 0.06014496460556984
step: 430, loss: 0.012611046433448792
step: 440, loss: 0.0015587212983518839
step: 450, loss: 0.012587265111505985
step: 460, loss: 0.019483372569084167
step: 470, loss: 0.058278173208236694
step: 480, loss: 0.005884575191885233
step: 490, loss: 0.0013040892081335187
step: 500, loss: 0.006963672116398811
step: 510, loss: 0.00015132312546484172
step: 520, loss: 0.021482836455106735
step: 530, loss: 0.00724908523261547
/usr/local/lib/python3.7/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=0.9386416861826699, f1=0.9347014925373135,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0029755395371466875
step: 10, loss: 0.0026004917453974485
step: 20, loss: 0.006453835871070623
step: 30, loss: 0.03736880421638489
step: 40, loss: 0.00018265216203872114
step: 50, loss: 0.00289634452201426603
step: 60, loss: 0.027432886883616447
step: 70, loss: 0.03425974398851395
step: 80, loss: 0.08957130461931229
step: 90, loss: 0.00016011190018616617
step: 100, loss: 0.0004105189291294664
step: 110, loss: 0.0006752859335392714
step: 120, loss: 9.771736222319305e-05
step: 130, loss: 0.00012763880658894777
step: 140, loss: 0.0019569932483136654
step: 150, loss: 0.00029959488892927766
step: 160, loss: 0.0018008442129939795
step: 170, loss: 0.004207056947052479
step: 180, loss: 0.0016023009084165096
step: 190, loss: 7.750474469503388e-05
step: 200, loss: 0.0007426406373269856
step: 210, loss: 0.00023152252833824605
step: 220, loss: 0.0003252573951613158
step: 230, loss: 0.0007006254163570702
step: 240, loss: 0.0026388554833829403
step: 250, loss: 0.0013650148175656796
step: 260, loss: 0.00632954528555274
step: 270, loss: 0.005565791390836239
step: 280, loss: 0.01899988390505314
step: 290, loss: 0.004067597910761833
step: 300, loss: 0.0022083609364926815
step: 310, loss: 0.01887970231473446
step: 320, loss: 0.02998882532119751
step: 330, loss: 0.0015949945664033294
step: 340, loss: 0.00220046681351959597
step: 350, loss: 0.012358097359538078
step: 360, loss: 0.0029554767534136677
step: 370, loss: 0.03997280076146126
step: 380, loss: 0.0042970688082277775
step: 390, loss: 0.00012206481915200129
step: 400, loss: 0.00025170185836032033
step: 410, loss: 0.0040782783180475235
step: 420, loss: 0.0006897789426147938
step: 430, loss: 0.0004951549344696105
step: 440, loss: 0.0001742251479299739
step: 450, loss: 0.0025301678106188774
```

```
step: 460, loss: 0.005987069103866816
step: 470, loss: 0.022058360278606415
step: 480, loss: 0.000872981094289571
step: 490, loss: 0.004454521462321281
step: 500, loss: 6.667841080343351e-05
step: 510, loss: 0.05320613086223602
step: 520, loss: 0.003190418239682913
step: 530, loss: 0.00012699983199127018
/usr/local/lib/python3.7/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=0.9402427637721755, f1=0.9365446966188051,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004743017256259918
step: 10, loss: 0.00309782521799920673
step: 20, loss: 0.003389465855434537
step: 30, loss: 0.012000731192529202
step: 40, loss: 0.001254621776752174
step: 50, loss: 0.00018712796736508608
step: 60, loss: 8.027870353544131e-05
step: 70, loss: 0.0008554699597880244
step: 80, loss: 0.001746101421304047
step: 90, loss: 0.0012856137473136187
step: 100, loss: 0.0024264713283628225
step: 110, loss: 0.029029546305537224
step: 120, loss: 0.0008103348081931472
step: 130, loss: 0.005029710941016674
step: 140, loss: 3.725942951859906e-05
step: 150, loss: 0.0012524735648185015
step: 160, loss: 0.006676336284726858
step: 170, loss: 0.002725715981796384
step: 180, loss: 0.00039380567613989115
step: 190, loss: 0.0006307719158940017
step: 200, loss: 0.0007637465605512261
step: 210, loss: 0.00025416878634132445
step: 220, loss: 0.0005042069242335856
step: 230, loss: 0.022594664245843887
step: 240, loss: 0.00920860841870308
step: 250, loss: 0.008351013995707035
step: 260, loss: 0.04481397196650505
step: 270, loss: 6.690831651212648e-05
step: 280, loss: 0.011528832837939262
step: 290, loss: 0.00046897243009880185
step: 300, loss: 0.0011701113100670278
step: 310, loss: 0.023572484031319618
```

```
step: 320, loss: 0.0014506358420476317
step: 330, loss: 0.024213097989559174
step: 340, loss: 0.0012094585690647364
step: 350, loss: 0.1193007379770279
step: 360, loss: 0.0039190929383039474
step: 370, loss: 0.0030190772376954556
step: 380, loss: 0.01964881271123886
step: 390, loss: 0.0002538417757023126
step: 400, loss: 0.036396343261003494
step: 410, loss: 0.0001978395157493651
step: 420, loss: 0.00017256260616704822
step: 430, loss: 0.001291221589781344
step: 440, loss: 0.0019884437788277864
step: 450, loss: 0.002397378906607628
step: 460, loss: 0.0007854824652895331
step: 470, loss: 0.007017569150775671
step: 480, loss: 6.927855429239571e-05
step: 490, loss: 0.0003918932343367487
step: 500, loss: 0.00010294202365912497
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 510, loss: 8.836261986289173e-05
step: 520, loss: 0.00027295894687995315
step: 530, loss: 0.003054923377931118
/usr/local/lib/python3.7/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=0.9421028253821214, f1=0.9425287356321841,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009530772222205997
step: 10, loss: 0.005012162495404482
step: 20, loss: 4.635108052752912e-05
step: 30, loss: 0.008678619749844074
step: 40, loss: 0.001847477862611413
step: 50, loss: 0.002649318426847458
step: 60, loss: 0.0003079960006289184
step: 70, loss: 0.07964812964200974
step: 80, loss: 0.0015427909092977643
step: 90, loss: 0.01902017369866371
step: 100, loss: 0.00016639960813336074
step: 110, loss: 0.0033645883668214083
step: 120, loss: 0.0013743912568315864
step: 130, loss: 0.20918409526348114
step: 140, loss: 0.030424179509282112
step: 150, loss: 0.003246043110266328
step: 160, loss: 0.00013593111361842602
```

```
step: 170, loss: 0.0004922394873574376
step: 180, loss: 0.00014191170339472592
step: 190, loss: 0.0012216907925903797
step: 200, loss: 0.003551325760781765
step: 210, loss: 0.00023717348813079298
step: 220, loss: 0.010106422007083893
step: 230, loss: 0.0026540574617683887
step: 240, loss: 0.000567532901186496
step: 250, loss: 0.0031048511154949665
step: 260, loss: 0.0004279104759916663
step: 270, loss: 0.00043563114013522863
step: 280, loss: 6.29359929007478e-05
step: 290, loss: 0.006818260531872511
step: 300, loss: 0.00015964066551532596
step: 310, loss: 0.00010977951751556247
step: 320, loss: 0.0024837187957018614
step: 330, loss: 0.0002557278494350612
step: 340, loss: 7.863562495913357e-05
step: 350, loss: 0.0009510628879070282
step: 360, loss: 0.012756857089698315
step: 370, loss: 0.00023572234204038978
step: 380, loss: 0.00011014382471330464
step: 390, loss: 0.006664849352091551
step: 400, loss: 0.10665348172187805
step: 410, loss: 0.00017116504022851586
step: 420, loss: 0.003073544707149267
step: 430, loss: 4.6458953875117004e-05
step: 440, loss: 0.0019217071821913123
step: 450, loss: 0.000268334784777835
step: 460, loss: 0.08761946111917496
step: 470, loss: 0.0446246899664402
step: 480, loss: 0.00015019639977253973
step: 490, loss: 0.00029965603607706726
step: 500, loss: 0.040642764419317245
step: 510, loss: 0.004961784463375807
step: 520, loss: 0.0008906465955078602
step: 530, loss: 0.00027914944803342223
/usr/local/lib/python3.7/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=0.9420491423273065, f1=0.9413394919168592,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003883334808051586
step: 10, loss: 0.000367186643416062
step: 20, loss: 0.0007914547459222376
```

```
step: 30, loss: 5.9976398915750906e-05
step: 40, loss: 0.0006418753182515502
step: 50, loss: 4.685541716753505e-05
step: 60, loss: 0.0004684669547714293
step: 70, loss: 0.00111811060924083
step: 80, loss: 0.00039718617335893214
step: 90, loss: 8.294724102597684e-05
step: 100, loss: 0.00014584460586775094
step: 110, loss: 0.002281059743836522
step: 120, loss: 0.00010327160998713225
step: 130, loss: 0.001374946441501379
step: 140, loss: 0.002037984551861882
step: 150, loss: 7.820752216503024e-05
step: 160, loss: 0.00010403939813841134
step: 170, loss: 0.03764490410685539
step: 180, loss: 0.019754674285650253
step: 190, loss: 0.00631311412900686
step: 200, loss: 0.003759943414479494
step: 210, loss: 0.000984518090263009
step: 220, loss: 0.0003741081163752824
step: 230, loss: 0.007928797975182533
step: 240, loss: 0.00017601462604943663
step: 250, loss: 0.0029968481976538897
step: 260, loss: 0.026458974927663803
step: 270, loss: 0.0013982668751850724
step: 280, loss: 0.000555817096028477
step: 290, loss: 0.0015130164101719856
step: 300, loss: 0.00798036903142929
step: 310, loss: 0.037775129079818726
step: 320, loss: 0.0008215152192860842
step: 330, loss: 0.0024051980581134558
step: 340, loss: 0.002524256007745862
step: 350, loss: 0.0014940939145162702
step: 360, loss: 0.00013639214739669114
step: 370, loss: 0.0028112968429923058
step: 380, loss: 0.005237694829702377
step: 390, loss: 5.507158857653849e-05
step: 400, loss: 0.00033836811780929565
step: 410, loss: 0.00016504013910889626
step: 420, loss: 0.00657090125605464
step: 430, loss: 0.00013543010572902858
step: 440, loss: 0.00016840627358760685
step: 450, loss: 0.02199966087937355
step: 460, loss: 0.005464180372655392
step: 470, loss: 0.00012209081614855677
step: 480, loss: 0.005945364944636822
step: 490, loss: 0.09532272070646286
step: 500, loss: 0.015495595522224903
```

```
step: 510, loss: 0.0006389273330569267
step: 520, loss: 0.015881305560469627
step: 530, loss: 6.305224815150723e-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=0.9448307834955958, f1=0.9383624655013799,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007759614381939173
step: 10, loss: 0.0010663828579708934
step: 20, loss: 0.0002484366996213794
step: 30, loss: 0.0009203495574183762
step: 40, loss: 0.00011134044325444847
step: 50, loss: 9.196899918606505e-05
step: 60, loss: 0.00017906395078171045
step: 70, loss: 2.7275518732494675e-05
step: 80, loss: 0.0002450529136694968
step: 90, loss: 0.0004588172014337033
step: 100, loss: 0.0029530164320021868
step: 110, loss: 0.003089603502303362
step: 120, loss: 0.012884038500487804
step: 130, loss: 0.00025978763005696237
step: 140, loss: 0.001433061552233994
step: 150, loss: 7.748894131509587e-05
step: 160, loss: 0.0009635738679207861
step: 170, loss: 0.0004849542456213385
step: 180, loss: 0.0012927871430292726
step: 190, loss: 0.00015678101044613868
step: 200, loss: 0.004591960459947586
step: 210, loss: 0.0007782323518767953
step: 220, loss: 0.04536323994398117
step: 230, loss: 0.00036811994505114853
step: 240, loss: 0.0002465003926772624
step: 250, loss: 0.0005450538010336459
step: 260, loss: 0.0005914647481404245
step: 270, loss: 0.008470514789223671
step: 280, loss: 0.00026662222808226943
step: 290, loss: 0.08242592215538025
step: 300, loss: 0.012592488899827003
step: 310, loss: 0.00018014518718700856
step: 320, loss: 0.023638416081666946
step: 330, loss: 0.000945265986956656
step: 340, loss: 0.0018498116405680776
step: 350, loss: 0.015415041707456112
step: 360, loss: 0.0007078800699673593
```

```
step: 370, loss: 2.8780646971426904e-05
step: 380, loss: 3.387677497812547e-05
step: 390, loss: 0.015553169883787632
step: 400, loss: 2.8464171919040382e-05
step: 410, loss: 6.367781315930188e-05
step: 420, loss: 0.0001787625951692462
step: 430, loss: 0.06708350777626038
step: 440, loss: 9.427966142538935e-05
step: 450, loss: 0.0003458688734099269
step: 460, loss: 0.02421996369957924
step: 470, loss: 0.00019003471243195236
step: 480, loss: 6.882475281599909e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 490, loss: 0.0025385781191289425
step: 500, loss: 0.0007348469807766378
step: 510, loss: 0.0005687017110176384
step: 520, loss: 6.286163261393085e-05
step: 530, loss: 0.00016201917605940253
/usr/local/lib/python3.7/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=0.9435370975268316, f1=0.9410125406409661,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023350045375991613
step: 10, loss: 0.008609440177679062
step: 20, loss: 0.017052218317985535
step: 30, loss: 0.0005848098080605268
step: 40, loss: 6.131965346867219e-05
step: 50, loss: 0.0009630626882426441
step: 60, loss: 0.2776528000831604
step: 70, loss: 0.005364206153899431
step: 80, loss: 0.008101149462163448
step: 90, loss: 0.005484085530042648
step: 100, loss: 0.00012519162555690855
step: 110, loss: 8.54953977977857e-05
step: 120, loss: 0.0002877720689866692
step: 130, loss: 0.00010272410145262256
step: 140, loss: 0.001560610020533204
step: 150, loss: 0.0005258064484223723
step: 160, loss: 0.0003060273302253357
step: 170, loss: 0.00015415935195051134
step: 180, loss: 0.0015190119156613946
step: 190, loss: 6.877570558572188e-05
step: 200, loss: 0.0002007195434998721
step: 210, loss: 0.00013897581084165722
```

```
step: 220, loss: 2.896729711210355e-05
step: 230, loss: 0.0010126340202987194
step: 240, loss: 0.00019946777320001274
step: 250, loss: 0.00022778674610890448
step: 260, loss: 0.001540757599286735
step: 270, loss: 0.00285550975240767
step: 280, loss: 0.003501231549307704
step: 290, loss: 0.0029222010634839535
step: 300, loss: 0.0004131766618229449
step: 310, loss: 0.0003127145173493773
step: 320, loss: 0.00031327211763709784
step: 330, loss: 5.14053062943276e-05
step: 340, loss: 0.0013988051796332002
step: 350, loss: 0.019516395404934883
step: 360, loss: 0.0012414301745593548
step: 370, loss: 0.00012600066838786006
step: 380, loss: 0.0010173942428082228
step: 390, loss: 0.0025585810653865337
step: 400, loss: 0.001140535343438387
step: 410, loss: 0.018544111400842667
step: 420, loss: 0.00023282470647245646
step: 430, loss: 0.0010363160399720073
step: 440, loss: 0.0011114862281829119
step: 450, loss: 0.0021031915675848722
step: 460, loss: 0.005520261358469725
step: 470, loss: 0.03600391000509262
step: 480, loss: 0.0004495218163356185
step: 490, loss: 0.02496533840894699
step: 500, loss: 0.006119758356362581
step: 510, loss: 0.0002632827963680029
step: 520, loss: 0.01958903856575489
step: 530, loss: 0.00030666415113955736
/usr/local/lib/python3.7/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=0.9411764705882353, f1=0.942271880819367,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00030733313178643584
step: 10, loss: 0.0007948133279569447
step: 20, loss: 0.025433950126171112
step: 30, loss: 0.00711237546056509
step: 40, loss: 0.0004709051863756031
step: 50, loss: 0.00019455060828477144
step: 60, loss: 0.00029737374279648066
step: 70, loss: 0.0003008959465660155
```

```
step: 80, loss: 5.360677096177824e-05
step: 90, loss: 0.0003893878310918808
step: 100, loss: 0.0009659012430347502
step: 110, loss: 0.001298112329095602
step: 120, loss: 0.045349977910518646
step: 130, loss: 0.0007355646812357008
step: 140, loss: 9.561012120684609e-05
step: 150, loss: 0.001478802296333015
step: 160, loss: 0.00012341365800239146
step: 170, loss: 0.0001757360005285591
step: 180, loss: 9.180449706036597e-05
step: 190, loss: 0.00216464139521122
step: 200, loss: 0.010549577884376049
step: 210, loss: 0.0015832511708140373
step: 220, loss: 0.005473601631820202
step: 230, loss: 0.003743255278095603
step: 240, loss: 0.0001177395970444195
step: 250, loss: 0.046883199363946915
step: 260, loss: 0.00018171302508562803
step: 270, loss: 0.001352975144982338
step: 280, loss: 0.0001097815838875249
step: 290, loss: 0.00015063291357364506
step: 300, loss: 0.0004263315349817276
step: 310, loss: 0.0007944436511024833
step: 320, loss: 0.04288212209939957
step: 330, loss: 0.00012403160508256406
step: 340, loss: 0.0004480103380046785
step: 350, loss: 0.020012307912111282
step: 360, loss: 5.37977357453201e-05
step: 370, loss: 0.0001192262934637256
step: 380, loss: 0.0006950795650482178
step: 390, loss: 0.000497685803566128
step: 400, loss: 0.0015702027594670653
step: 410, loss: 0.00014773098519071937
step: 420, loss: 0.0002577516424935311
step: 430, loss: 0.00026369455736130476
step: 440, loss: 0.00019339284335728735
step: 450, loss: 0.00013737892732024193
step: 460, loss: 2.9313217964954674e-05
step: 470, loss: 0.0378994382917881
step: 480, loss: 0.12491550296545029
step: 490, loss: 0.00011862659448524937
step: 500, loss: 0.0019485853845253587
step: 510, loss: 0.0007583744009025395
step: 520, loss: 5.2313720516394824e-05
step: 530, loss: 0.0001297193084610626
/usr/local/lib/python3.7/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=0.9426038264115725, f1=0.9425393883225208,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.11785362707451e-05
step: 10, loss: 0.00023594086815137416
step: 20, loss: 7.688314508413896e-05
step: 30, loss: 0.00023627650807611644
step: 40, loss: 0.004341877065598965
step: 50, loss: 0.0014183592284098268
step: 60, loss: 8.04059236543253e-05
step: 70, loss: 6.880741420900449e-05
step: 80, loss: 0.000532486301381141
step: 90, loss: 0.00017809480777941644
step: 100, loss: 0.00022284052101895213
step: 110, loss: 2.8017255317536183e-05
step: 120, loss: 0.0001533535250928253
step: 130, loss: 0.0005152272642590106
step: 140, loss: 0.027424804866313934
step: 150, loss: 0.00015671708388254046
step: 160, loss: 0.0008217096328735352
step: 170, loss: 0.0033508813939988613
step: 180, loss: 0.0001801305916160345
step: 190, loss: 6.6921413235832e-05
step: 200, loss: 8.685527427587658e-05
step: 210, loss: 0.006116168573498726
step: 220, loss: 0.00016006632358767092
step: 230, loss: 0.0008041805122047663
step: 240, loss: 0.001037777867168188
step: 250, loss: 0.10507848858833313
step: 260, loss: 2.7667105314321816e-05
step: 270, loss: 0.00254929531365633
step: 280, loss: 0.00012336722284089774
step: 290, loss: 0.00041730329394340515
step: 300, loss: 0.0007251250208355486
step: 310, loss: 0.015673119574785233
step: 320, loss: 0.0004852578858844936
step: 330, loss: 0.00020843118545599282
step: 340, loss: 3.6949430068489164e-05
step: 350, loss: 0.004082266241312027
step: 360, loss: 0.002283574780449271
step: 370, loss: 0.00035459993523545563
step: 380, loss: 6.391799252014607e-05
step: 390, loss: 8.148645429173484e-05
step: 400, loss: 5.214736665948294e-05
step: 410, loss: 0.003656546352431178
```

```
step: 420, loss: 0.00043528134119696915
step: 430, loss: 2.991680048580747e-05
step: 440, loss: 2.4031372959143482e-05
step: 450, loss: 6.282456160988659e-05
step: 460, loss: 0.00202928320504725
step: 470, loss: 0.000395460199797526
step: 480, loss: 5.161493390914984e-05
step: 490, loss: 0.00013659482647199184
step: 500, loss: 0.0006799664115533233
step: 510, loss: 0.00018528776126913726
step: 520, loss: 0.00019242837151978165
step: 530, loss: 0.0012844082666561007
/usr/local/lib/python3.7/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=0.9422180801491147, f1=0.9459584295612009,
best_f1=0.9344188150158298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.506906644674018e-05
step: 10, loss: 0.00016347484779544175
step: 20, loss: 0.002618790604174137
step: 30, loss: 0.0005179444560781121
step: 40, loss: 6.331441545626149e-05
step: 50, loss: 0.00048708103713579476
step: 60, loss: 4.772823740495369e-05
step: 70, loss: 0.00012455815158318728
step: 80, loss: 0.00013524301175493747
step: 90, loss: 6.552047852892429e-05
step: 100, loss: 0.00013987446436658502
step: 110, loss: 0.021624602377414703
step: 120, loss: 0.0010077640181407332
step: 130, loss: 5.563842933042906e-05
step: 140, loss: 6.322530680336058e-05
step: 150, loss: 0.0001541399396955967
step: 160, loss: 0.00013394666893873364
step: 170, loss: 7.955613546073437e-05
step: 180, loss: 0.000153328335727565
step: 190, loss: 0.00015348067972809076
step: 200, loss: 0.00018711763550527394
step: 210, loss: 0.00011150001228088513
step: 220, loss: 9.170162957161665e-05
step: 230, loss: 3.766049849218689e-05
step: 240, loss: 0.0005265119834803045
step: 250, loss: 0.0003271101741120219
step: 260, loss: 0.0005171149387024343
step: 270, loss: 0.05823814868927002
```

```
step: 280, loss: 0.0007171790930442512
step: 290, loss: 0.0005392606253735721
step: 300, loss: 0.001525334664620459
step: 310, loss: 0.0012048223288729787
step: 320, loss: 0.00021118979202583432
step: 330, loss: 0.0004456697206478566
step: 340, loss: 0.0010729492641985416
step: 350, loss: 0.0003633481392171234
step: 360, loss: 0.00022381439339369535
step: 370, loss: 3.5042379749938846e-05
step: 380, loss: 0.0008412235183641315
step: 390, loss: 0.0003721341781783849
step: 400, loss: 0.00013486907118931413
step: 410, loss: 0.0006834790110588074
step: 420, loss: 5.867647269042209e-05
step: 430, loss: 4.064239328727126e-05
step: 440, loss: 0.007891412824392319
step: 450, loss: 0.0033780436497181654
step: 460, loss: 0.0001696286053629592
step: 470, loss: 0.00016050811973400414
step: 480, loss: 0.0006836397806182504
step: 490, loss: 0.002514231950044632
step: 500, loss: 0.0003507660876493901
step: 510, loss: 0.003069723956286907
step: 520, loss: 4.407859159982763e-05
step: 530, loss: 3.52312381437514e-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=0.9412313432835822, f1=0.9454209065679925,
best_f1=0.9344188150158298
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
```

```
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:16, 342.57it/s]
load_f1 = 0.9428962996802193
real_f1 = 0.9397260273972603
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, 377.90it/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
```

```python
# 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 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: 0.8616650700569153

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.288659793814433, f1=0.2692307692307693,
best_f1=0.2692307692307693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3710891604423523
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.2857142857142857, f1=0.4666666666666667,
best_f1=0.2692307692307693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3122541904449463
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.32941176470588235, f1=0.32941176470588235,
best_f1=0.32941176470588235
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3652125298976898
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.3902439024390244, f1=0.3548387096774194,
best_f1=0.3548387096774194
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.27422380447387695
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4210526315789474, f1=0.32432432432432434,
best_f1=0.32432432432432434
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23717041313648224
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 6: dev_f1=0.48888888888888893, f1=0.3582089552238806,
best_f1=0.3582089552238806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1745259314775467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5581395348837208, f1=0.3880597014925373,
best_f1=0.3880597014925373
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20249563455581665
/usr/local/lib/python3.7/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=0.5517241379310344, f1=0.4878048780487805,
best_f1=0.3880597014925373
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.09756321460008621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5806451612903226, f1=0.45, best_f1=0.45
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08327936381101608
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6000000000000001, 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: 0.056672438979148865
/usr/local/lib/python3.7/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=0.6000000000000001, 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: 0.12233240157365799
/usr/local/lib/python3.7/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=0.6000000000000001, f1=0.45, 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: 0.024198459461331367
/usr/local/lib/python3.7/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=0.6000000000000001, f1=0.5, 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: 0.07959699630737305
/usr/local/lib/python3.7/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=0.6000000000000001, f1=0.5, 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: 0.12049926817417145
/usr/local/lib/python3.7/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=0.6000000000000001, f1=0.5, best_f1=0.4615384615384615
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.")
91it [00:00, 127823.73it/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 = 0.6000000000000001
real_f1 = 0.5454545454545454
```

```
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 438.82it/s]
```

2.1.2 DBLP-ACM - Running the matcher

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

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

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

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

Defaults for this optimization level are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
```

140

```
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: 0.8116663694381714
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.47489166259765625
step: 20, loss: 0.5896987915039062
step: 30, loss: 0.44652456045150757
step: 40, loss: 0.2072359174489975
step: 50, loss: 0.057666923850774765
step: 60, loss: 0.11024200916290283
step: 70, loss: 0.05601911246776581
step: 80, loss: 0.13449153304100037
step: 90, loss: 0.00688742334023118
step: 100, loss: 0.01122188474982977
step: 110, loss: 0.09280744940042496
step: 120, loss: 0.04110274463891983
step: 130, loss: 0.005102312657982111
step: 140, loss: 0.07019738107919693
step: 150, loss: 0.13342590630054474
step: 160, loss: 0.12271621823310852
step: 170, loss: 0.011602882295846939
step: 180, loss: 0.04750337079167366
step: 190, loss: 0.012884972617030144
step: 200, loss: 0.004298060666769743
step: 210, loss: 0.0393809899687767
```

```
step: 220, loss: 0.0037183149252086878
step: 230, loss: 0.004057024605572224
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9865470852017937, f1=0.9876543209876544,
best_f1=0.9876543209876544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007435173261910677
step: 10, loss: 0.001180833438411355
step: 20, loss: 0.0026007541455328465
step: 30, loss: 0.0014494986971840262
step: 40, loss: 0.0077840364538133144
step: 50, loss: 0.002613153774291277
step: 60, loss: 0.0016852879198268056
step: 70, loss: 0.004456221126019955
step: 80, loss: 0.0007307771011255682
step: 90, loss: 0.004604995716363192
step: 100, loss: 0.007403767202049494
step: 110, loss: 0.1529739499092102
step: 120, loss: 0.0015333298360928893
step: 130, loss: 0.002572824480012059
step: 140, loss: 0.009539413265883923
step: 150, loss: 0.00809943862259388
step: 160, loss: 0.03434740751981735
step: 170, loss: 0.07243844121694565
step: 180, loss: 0.0038100748788565397
step: 190, loss: 0.2257014513015747
step: 200, loss: 0.0028373009990900755
step: 210, loss: 0.15508708357810974
step: 220, loss: 0.0018189754337072372
step: 230, loss: 0.1475621908903122
/usr/local/lib/python3.7/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=0.9809203142536477, f1=0.9774266365688488,
best_f1=0.9876543209876544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11533623188734055
step: 10, loss: 0.0031400583684444427
```

```
step: 20, loss: 0.010846509598195553
step: 30, loss: 0.00946712214499712
step: 40, loss: 0.009887956082820892
step: 50, loss: 0.007010126486420631
step: 60, loss: 0.0012457198463380337
step: 70, loss: 0.002116021467372775
step: 80, loss: 0.0026683262549340725
step: 90, loss: 0.001097862608730793
step: 100, loss: 0.0010319709545001388
step: 110, loss: 0.014556600712239742
step: 120, loss: 0.000812107406090945
step: 130, loss: 0.000541751622222364
step: 140, loss: 0.0016314013628289104
step: 150, loss: 0.0019910342525690794
step: 160, loss: 0.001260695280507207
step: 170, loss: 0.0037432548124343157
step: 180, loss: 0.007166250608861446
step: 190, loss: 0.009966058656573296
step: 200, loss: 0.00096152164041996
step: 210, loss: 0.0156742874532938
step: 220, loss: 0.024283526465296745
step: 230, loss: 0.016482388600707054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9909706546275394, f1=0.9876265466816648,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005060170311480761
step: 10, loss: 0.0006027831113897264
step: 20, loss: 0.0003845828177873045
step: 30, loss: 0.0008844544645398855
step: 40, loss: 0.002338590333238244
step: 50, loss: 0.004147726111114025
step: 60, loss: 0.0005217368598096073
step: 70, loss: 0.03297387808561325
step: 80, loss: 0.04781671240925789
step: 90, loss: 0.023428289219737053
step: 100, loss: 0.0008988314075395465
step: 110, loss: 0.016862893477082253
step: 120, loss: 0.0078784404322505
step: 130, loss: 0.0045201657339930534
```

```
step: 140, loss: 0.0003132162964902818
step: 150, loss: 0.0036494573578238487
step: 160, loss: 0.0012027234770357609
step: 170, loss: 0.017680631950497627
step: 180, loss: 0.03565552458167076
step: 190, loss: 0.0019371655071154237
step: 200, loss: 0.0004767997015733272
step: 210, loss: 0.0005132638034410775
step: 220, loss: 0.00043520500184968114
step: 230, loss: 0.0006295783095993102
```
/usr/local/lib/python3.7/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=0.9899216125419933, f1=0.980963045912654,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```
step: 0, loss: 0.0005713690188713372
step: 10, loss: 0.004165814258158207
step: 20, loss: 0.00035263790050521493
step: 30, loss: 0.00017797511827666312
step: 40, loss: 0.0004484712262637913
step: 50, loss: 0.0004941513179801404
step: 60, loss: 0.0004275232495274395
step: 70, loss: 0.00022992234153207392
step: 80, loss: 0.001072409562766552
step: 90, loss: 0.000903724052477628
step: 100, loss: 0.003124253125861287
step: 110, loss: 0.0014978173421695828
step: 120, loss: 0.08903991430997849
step: 130, loss: 0.067832261322392883
step: 140, loss: 0.0004658429534174502
step: 150, loss: 0.00020711115159196779
step: 160, loss: 0.0023695826530456543
step: 170, loss: 0.025827016681432724
step: 180, loss: 0.0005285727675072849
step: 190, loss: 0.0009166765375994146
step: 200, loss: 0.0038234745152294636
step: 210, loss: 0.0006181607022881508
step: 220, loss: 0.0005408124998211861
step: 230, loss: 0.0005301738856360316
```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9921436588103255, f1=0.9876265466816648,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006813303334638476
step: 10, loss: 0.008393609896302223
step: 20, loss: 0.0006957122241146863
step: 30, loss: 0.00019724674348253757
step: 40, loss: 0.0005189247895032167
step: 50, loss: 0.004412796348333359
step: 60, loss: 0.014932097867131233
step: 70, loss: 0.0005600089789368212
step: 80, loss: 0.0003418914566282183
step: 90, loss: 0.00026544296997599304
step: 100, loss: 0.000457530579296872
step: 110, loss: 0.029508551582694054
step: 120, loss: 0.0007351007661782205
step: 130, loss: 0.0009330059983767569
step: 140, loss: 0.02465238608419895
step: 150, loss: 0.0005044665886089206
step: 160, loss: 0.009316584095358849
step: 170, loss: 0.0005223533371463418
step: 180, loss: 0.0004714417736977339
step: 190, loss: 0.0008928694878704846
step: 200, loss: 0.021643076092004776
step: 210, loss: 0.0010060496861115098
step: 220, loss: 0.0006015229737386107
step: 230, loss: 0.00027462412253953516
/usr/local/lib/python3.7/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=0.9921259842519685, f1=0.984304932735426,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.044533729553222656
step: 10, loss: 0.0001575742062414065
step: 20, loss: 0.00032387845567427576
step: 30, loss: 0.008926165290176868
step: 40, loss: 0.0002874610072467476
step: 50, loss: 0.00010233387729385868
step: 60, loss: 0.0023221734445542097
step: 70, loss: 0.00018841946439351887
step: 80, loss: 0.00013864068023394793
step: 90, loss: 0.00020549762120936066
```

```
step: 100, loss: 0.0006783364806324244
step: 110, loss: 0.022385932505130768
step: 120, loss: 0.001061460468918085
step: 130, loss: 0.0005367104895412922
step: 140, loss: 0.0001903451920952648
step: 150, loss: 0.00021853383805137128
step: 160, loss: 0.0002557039842940867
step: 170, loss: 0.00023920342209748924
step: 180, loss: 0.0017164768651127815
step: 190, loss: 0.012091618031263351
step: 200, loss: 0.0005306469975039363
step: 210, loss: 0.00017797420150600374
step: 220, loss: 0.0004450898850336671
step: 230, loss: 0.1984919160604477
/usr/local/lib/python3.7/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=0.987736900780379, f1=0.9821826280623607,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07976393401622772
step: 10, loss: 0.019840989261865616
step: 20, loss: 0.0010965579422190785
step: 30, loss: 0.013022989965975285
step: 40, loss: 0.005964596755802631
step: 50, loss: 0.004030171781778336
step: 60, loss: 0.0002626122732181102
step: 70, loss: 0.0021181993652135134
step: 80, loss: 0.0008992849034257233
step: 90, loss: 0.0004638547252397984
step: 100, loss: 0.0002244489878648892
step: 110, loss: 0.0005757449544034898
step: 120, loss: 0.00025834949337877333
step: 130, loss: 0.18184368312358856
step: 140, loss: 0.0001867717510322109
step: 150, loss: 0.0009722657268866897
step: 160, loss: 0.0003133649006485939
step: 170, loss: 0.002656431868672371
step: 180, loss: 0.0005574904498644173
step: 190, loss: 0.00036790629383176565
step: 200, loss: 0.001002630335278809
step: 210, loss: 0.000323600834235549
step: 220, loss: 0.0005564085440710187
step: 230, loss: 0.02331618219614029
/usr/local/lib/python3.7/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=0.987736900780379, f1=0.9800443458980044,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022545228421222419
step: 10, loss: 0.0002507358731236309
step: 20, loss: 0.0008658675360493362
step: 30, loss: 0.00019169668667018414
step: 40, loss: 0.00018739895313046873
step: 50, loss: 0.00022576414630748332
step: 60, loss: 0.0003258686338085681
step: 70, loss: 0.000434497109381482
step: 80, loss: 0.035858187824487686
step: 90, loss: 0.000730806787032634
step: 100, loss: 0.0001806330110412091
step: 110, loss: 0.00010587471479084343
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 120, loss: 0.04259597510099411
step: 130, loss: 0.00012114741548430175
step: 140, loss: 0.00013169091835152358
step: 150, loss: 0.00014593418745789677
step: 160, loss: 0.00011681163596222177
step: 170, loss: 8.562720904592425e-05
step: 180, loss: 0.0001502670784248039
step: 190, loss: 6.516411667689681e-05
step: 200, loss: 0.0001281879813177511
step: 210, loss: 0.0005873282207176089
step: 220, loss: 0.03134353458881378
step: 230, loss: 0.012287004850804806
/usr/local/lib/python3.7/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=0.9898534385569334, f1=0.9832026875699889,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012336882355157286
step: 10, loss: 9.347280138172209e-05
step: 20, loss: 0.00011671060929074883
step: 30, loss: 0.00023603775480296463
step: 40, loss: 5.652768231811933e-05
step: 50, loss: 0.00026198080740869045
step: 60, loss: 0.035210274159908295
step: 70, loss: 0.00010788768122438341
step: 80, loss: 0.0006045120535418391
```

```
step: 90, loss: 0.00012226718536112458
step: 100, loss: 0.0006386361201293766
step: 110, loss: 0.0004639519320335239
step: 120, loss: 0.01042565144598484
step: 130, loss: 0.0018283112440258265
step: 140, loss: 0.02368680015206337
step: 150, loss: 0.0047124167904257774
step: 160, loss: 0.00011155172251164913
step: 170, loss: 0.00023226124176289886
step: 180, loss: 0.00045322603546082973
step: 190, loss: 0.0003300530952401459
step: 200, loss: 8.301081834360957e-05
step: 210, loss: 0.00012177408643765375
step: 220, loss: 0.00016165773558896035
step: 230, loss: 0.00019554924801923335
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.987598647125141, f1=0.9876265466816648,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011856589844683185
step: 10, loss: 0.0007875387091189623
step: 20, loss: 0.0017452476313337684
step: 30, loss: 0.000713892572093755
step: 40, loss: 0.00016585348930675536
step: 50, loss: 0.0007568986620754004
step: 60, loss: 0.0003400285786483437
step: 70, loss: 0.0005132585065439343
step: 80, loss: 0.0001977801730390638
step: 90, loss: 0.0008318166364915669
step: 100, loss: 0.00017454942280892283
step: 110, loss: 0.00014616618864238262
step: 120, loss: 0.00021140932221896946
step: 130, loss: 7.81910348450765e-05
step: 140, loss: 8.524220174876973e-05
step: 150, loss: 8.419041841989383e-05
step: 160, loss: 0.00020971798221580684
step: 170, loss: 0.02474764548242092
step: 180, loss: 0.0002378437202423811
step: 190, loss: 0.0003444281464908272
step: 200, loss: 0.00015205905947368592
step: 210, loss: 0.00034729408798739314
step: 220, loss: 7.382988405879587e-05
step: 230, loss: 0.011697111651301384
/usr/local/lib/python3.7/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=0.9866071428571428, f1=0.9778270509977827,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010222307173535228
step: 10, loss: 0.0015230559511110187
step: 20, loss: 0.00010469643893884495
step: 30, loss: 0.005361204035580158
step: 40, loss: 0.00016350572695955634
step: 50, loss: 0.00012064790644217283
step: 60, loss: 0.011708153411746025
step: 70, loss: 8.392820745939389e-05
step: 80, loss: 7.282487058546394e-05
step: 90, loss: 0.0001330416853306815
step: 100, loss: 0.00010907971591223031
step: 110, loss: 0.00011150785101344809
step: 120, loss: 0.0003159870975650847
step: 130, loss: 9.765819413587451e-05
step: 140, loss: 0.00012303033145144582
step: 150, loss: 0.00018983360496349633
step: 160, loss: 0.027285009622573853
step: 170, loss: 8.686412911629304e-05
step: 180, loss: 6.980084435781464e-05
step: 190, loss: 8.32053046906367e-05
step: 200, loss: 0.007756824605166912
step: 210, loss: 6.592108547920361e-05
step: 220, loss: 0.040766164660453796
step: 230, loss: 0.00010941179061774164
/usr/local/lib/python3.7/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=0.9899441340782122, f1=0.98, best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011156823893543333
step: 10, loss: 8.822737436275929e-05
step: 20, loss: 0.00023995839001145214
step: 30, loss: 0.03169528767466545
step: 40, loss: 0.0028780940920114517
step: 50, loss: 0.00013188783486839384
step: 60, loss: 7.624316640431061e-05
step: 70, loss: 0.00010127443238161504
step: 80, loss: 7.622106932103634e-05
step: 90, loss: 5.572690861299634e-05
```

```
step: 100, loss: 0.0003380195703357458
step: 110, loss: 0.00010374067642260343
step: 120, loss: 0.00016421084001194686
step: 130, loss: 0.00012591214908775259
step: 140, loss: 0.0016740558203309774
step: 150, loss: 0.01408927608281374
step: 160, loss: 7.212324999272823e-05
step: 170, loss: 8.398017234867439e-05
step: 180, loss: 0.00010348926298320293
step: 190, loss: 5.619552393909544e-05
step: 200, loss: 5.930128099862486e-05
step: 210, loss: 0.00013074334128759801
step: 220, loss: 0.00014151257346384227
step: 230, loss: 5.020925891585648e-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=0.978912319644839,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010492466390132904
step: 10, loss: 3.725482019945048e-05
step: 20, loss: 0.011816662736237049
step: 30, loss: 0.00017518557433350461
step: 40, loss: 4.220124537823722e-05
step: 50, loss: 7.627002923982218e-05
step: 60, loss: 0.00014887990255374461
step: 70, loss: 7.384435593849048e-05
step: 80, loss: 9.450947982259095e-05
step: 90, loss: 9.41946855164133e-05
step: 100, loss: 0.008287512697279453
step: 110, loss: 0.00012685437104664743
step: 120, loss: 0.00020097155356779695
step: 130, loss: 0.00013546450645662844
step: 140, loss: 0.00016966003749985248
step: 150, loss: 0.00013575034972745925
step: 160, loss: 0.0004706278268713504
step: 170, loss: 7.77912064222619e-05
step: 180, loss: 0.0001332326210103929
step: 190, loss: 0.00015906646149232984
step: 200, loss: 4.984154656995088e-05
step: 210, loss: 0.012638594955205917
step: 220, loss: 7.502056541852653e-05
step: 230, loss: 3.500056845950894e-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=0.9899665551839464, f1=0.9778270509977827,
best_f1=0.9876265466816648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.382560993894003e-05
step: 10, loss: 0.0009401098359376192
step: 20, loss: 7.052836008369923e-05
step: 30, loss: 0.0001462360960431397
step: 40, loss: 0.00011875099153257906
step: 50, loss: 0.00011830576841020957
step: 60, loss: 4.576592255034484e-05
step: 70, loss: 7.885850936872885e-05
step: 80, loss: 0.022703789174556732
step: 90, loss: 3.614115121308714e-05
step: 100, loss: 7.526991976192221e-05
step: 110, loss: 9.691231389297172e-05
step: 120, loss: 0.00012153519492130727
step: 130, loss: 0.00010647097224136814
step: 140, loss: 0.0001241448480868712
step: 150, loss: 9.555325232213363e-05
step: 160, loss: 0.0002841743698809296
step: 170, loss: 4.3531214032555e-05
step: 180, loss: 6.434885290218517e-05
step: 190, loss: 0.019322171807289124
step: 200, loss: 6.696590571664274e-05
step: 210, loss: 0.01978055015206337
step: 220, loss: 0.00012970756506547332
step: 230, loss: 0.00014583133452106267
/usr/local/lib/python3.7/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=0.9910514541387023, f1=0.978912319644839,
best_f1=0.9876265466816648
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
```

```
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:06, 361.51it/s]
load_f1 = 0.9932279909706545
real_f1 = 0.992108229988726
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 438.02it/s]
```

2.1.3 DBLP-GoogleScholar - Running the matcher

```python
[ ]: # 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
```

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

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

Defaults for this optimization level are:
enabled   : True
opt_level   : 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: 0.7914389967918396

```
step: 10, loss: 0.4279686212539673
step: 20, loss: 0.47869083285331726
step: 30, loss: 0.4020833373069763
step: 40, loss: 0.2961425185203552
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.15036877989768982
step: 60, loss: 0.15623410046100616
step: 70, loss: 0.22379232943058014
step: 80, loss: 0.07218462228775024
step: 90, loss: 0.09830053150653839
step: 100, loss: 0.32798609137535095
step: 110, loss: 0.0772087350487709
step: 120, loss: 0.04354904219508171
step: 130, loss: 0.014917982742190361
step: 140, loss: 0.25548654794692993
step: 150, loss: 0.07097642868757248
step: 160, loss: 0.19178485870361328
step: 170, loss: 0.2685132324695587
step: 180, loss: 0.06850049644708633
step: 190, loss: 0.07710840553045273
step: 200, loss: 0.11751306802034378
step: 210, loss: 0.10084813088178635
step: 220, loss: 0.10205379873514175
step: 230, loss: 0.0806652307510376
step: 240, loss: 0.040016621351242065
step: 250, loss: 0.05398146063089371
step: 260, loss: 0.02382315695285797
step: 270, loss: 0.01446917001157999
step: 280, loss: 0.06175648048520088
step: 290, loss: 0.04868099093437195
step: 300, loss: 0.08884581178426743
step: 310, loss: 0.09122951328754425
step: 320, loss: 0.0682646632194519
step: 330, loss: 0.11455991119146347
step: 340, loss: 0.17426921427249908
step: 350, loss: 0.07255580276250839
step: 360, loss: 0.08338931202888489
step: 370, loss: 0.10134592652320862
step: 380, loss: 0.16331657767295837
step: 390, loss: 0.01654195785522461
step: 400, loss: 0.005066131241619587
step: 410, loss: 0.006626848131418228
step: 420, loss: 0.0070093716494739056
step: 430, loss: 0.10117793083190918
step: 440, loss: 0.05219335854053497
step: 450, loss: 0.0036949063651263714
step: 460, loss: 0.2541431784629822
step: 470, loss: 0.2944287955760956
```

```
step: 480, loss: 0.22471311688423157
step: 490, loss: 0.043733205646276474
step: 500, loss: 0.00820870604366064
step: 510, loss: 0.07984592020511627
step: 520, loss: 0.022236758843064308
step: 530, loss: 0.15599416196346283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9378166743436204, f1=0.9357798165137614,
best_f1=0.9357798165137614
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05784078314900398
step: 10, loss: 0.109748750925506409
step: 20, loss: 0.10185786336660385
step: 30, loss: 0.043176136910915375
step: 40, loss: 0.007790958043187857
step: 50, loss: 0.08533140271902084
step: 60, loss: 0.2709912359714508
step: 70, loss: 0.06508022546768188
step: 80, loss: 0.01780470833182335
step: 90, loss: 0.00784648023545742
step: 100, loss: 0.39806845784187317
step: 110, loss: 0.030633896589279175
step: 120, loss: 0.07125160098075867
step: 130, loss: 0.020506111904978752
step: 140, loss: 0.029379069805145264
step: 150, loss: 0.05172418802976608
step: 160, loss: 0.04619349539279938
step: 170, loss: 0.1117679700255394
step: 180, loss: 0.0038288759533315897
step: 190, loss: 0.002303373534232378
step: 200, loss: 0.036131542176008224
step: 210, loss: 0.0014904828276485205
step: 220, loss: 0.18389815092086792
step: 230, loss: 0.07228948920965195
step: 240, loss: 0.12821049988269806
step: 250, loss: 0.08621380478143692
step: 260, loss: 0.007724678609520197
step: 270, loss: 0.137978658080101
step: 280, loss: 0.2625114619731903
step: 290, loss: 0.11355194449424744
```

```
step: 300, loss: 0.03013598918914795
step: 310, loss: 0.024862533435225487
step: 320, loss: 0.08069115877151489
step: 330, loss: 0.019578084349632263
step: 340, loss: 0.02785339578986168
step: 350, loss: 0.030355850234627724
step: 360, loss: 0.01141335815191269
step: 370, loss: 0.028069809079170227
step: 380, loss: 0.06496423482894897
step: 390, loss: 0.024907490238547325
step: 400, loss: 0.0379333533346653
step: 410, loss: 0.0023101461119949818
step: 420, loss: 0.003501101629808545
step: 430, loss: 0.017398569732904434
step: 440, loss: 0.012985399924218655
step: 450, loss: 0.012619296088814735
step: 460, loss: 0.10319460928440094
step: 470, loss: 0.014624311588704586
step: 480, loss: 0.26034000515937805
step: 490, loss: 0.035901181399822235
step: 500, loss: 0.06191489100456238
step: 510, loss: 0.07984893769025803
step: 520, loss: 0.1448642611503601
step: 530, loss: 0.1391221582889557
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9467345993515517, f1=0.9422632794457274,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0047297417186200062
step: 10, loss: 0.05892416089773178
step: 20, loss: 0.09708702564239502
step: 30, loss: 0.37814798951148987
step: 40, loss: 0.005289576016366482
step: 50, loss: 0.007221278268843889
step: 60, loss: 0.01789570413529873
step: 70, loss: 0.05299828201532364
step: 80, loss: 0.0029574530199170113
step: 90, loss: 0.016958434134721756
step: 100, loss: 0.0031851958483457565
step: 110, loss: 0.052110567688941956
```

```
step: 120, loss: 0.009404422715306282
step: 130, loss: 0.016809571534395218
step: 140, loss: 0.033870015293359756
step: 150, loss: 0.005504182074218988
step: 160, loss: 0.01299291756004095
step: 170, loss: 0.052453890442848206
step: 180, loss: 0.014467675238847733
step: 190, loss: 0.009642082266509533
step: 200, loss: 0.0028816726990044117
step: 210, loss: 0.004239638336002827
step: 220, loss: 0.010583274066448212
step: 230, loss: 0.05742456763982773
step: 240, loss: 0.007812134921550751
step: 250, loss: 0.0011413423344492912
step: 260, loss: 0.0004586254945024848
step: 270, loss: 0.0003919187583960593
step: 280, loss: 0.014104284346103668
step: 290, loss: 0.07523084431886673
step: 300, loss: 0.013035986572504044
step: 310, loss: 0.09637689590454102
step: 320, loss: 0.0675685927271843
step: 330, loss: 0.0022351858206009331
step: 340, loss: 0.0010142731480300426
step: 350, loss: 0.024510584771633148
step: 360, loss: 0.05080061033368111
step: 370, loss: 0.003389514284208417
step: 380, loss: 0.028435982763767242
step: 390, loss: 0.02636193484067917
step: 400, loss: 0.03697546571493149
step: 410, loss: 0.019427945837378502
step: 420, loss: 0.060918644070625305
step: 430, loss: 0.023115089163184166
step: 440, loss: 0.01943560130894184
step: 450, loss: 0.02086428552865982
step: 460, loss: 0.1414175182580948
step: 470, loss: 0.012282001785933971
step: 480, loss: 0.02572053112089634
step: 490, loss: 0.011941168457269669
step: 500, loss: 0.03670680895447731
step: 510, loss: 0.00825444608926773
step: 520, loss: 0.01730579510331154
step: 530, loss: 0.045319776982069016
/usr/local/lib/python3.7/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=0.9393090569561158, f1=0.9428704133766836,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009522989392280579
step: 10, loss: 0.010075679048895836
step: 20, loss: 0.0414431057870388
step: 30, loss: 0.00568476552143693
step: 40, loss: 0.0020189927890896797
step: 50, loss: 0.002164301695302129
step: 60, loss: 0.00027075500111095607
step: 70, loss: 0.003451884724199772
step: 80, loss: 0.09241241961717606
step: 90, loss: 0.057857997715473175
step: 100, loss: 0.0027311076410114765
step: 110, loss: 0.01006810087710619
step: 120, loss: 0.0015305698616430163
step: 130, loss: 0.08706455677747726
step: 140, loss: 0.10334637761116028
step: 150, loss: 0.0005600545555353165
step: 160, loss: 0.011068764142692089
step: 170, loss: 0.005051273852586746
step: 180, loss: 0.004118641372770071
step: 190, loss: 0.000342955143423751
step: 200, loss: 0.0004019574844278395
step: 210, loss: 0.00241860537789762
step: 220, loss: 0.0007611993351019919
step: 230, loss: 0.33495190739631653
step: 240, loss: 0.04216066747903824
step: 250, loss: 0.0017231156816706061
step: 260, loss: 0.10234801471233368
step: 270, loss: 0.006797280162572861
step: 280, loss: 0.0006390096386894584
step: 290, loss: 0.08882569521665573
step: 300, loss: 0.016129638999700546
step: 310, loss: 0.010217450559139252
step: 320, loss: 0.008726369589567184
step: 330, loss: 0.004563470836728811
step: 340, loss: 0.008869314566254616
step: 350, loss: 0.0005275837029330432
step: 360, loss: 0.009089292027056217
step: 370, loss: 0.001852476503700018
step: 380, loss: 0.0010939333587884903
step: 390, loss: 0.06328441947698593
step: 400, loss: 0.06773647665977478
step: 410, loss: 0.002761394251137972
step: 420, loss: 0.0017190471990033984
step: 430, loss: 0.00202531018294394
step: 440, loss: 0.10310222208499908
step: 450, loss: 0.02431228570640087
```

```
step: 460, loss: 0.00022576759511139244
step: 470, loss: 0.01153518445789814
step: 480, loss: 0.0023899420630186796
step: 490, loss: 0.007881168276071548
step: 500, loss: 0.0056315334513783455
step: 510, loss: 0.01616939902305603
step: 520, loss: 0.0006289507145993412
step: 530, loss: 0.0029062603134661913
/usr/local/lib/python3.7/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=0.9379770992366412, f1=0.9360687022900763,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001885063829831779
step: 10, loss: 0.0004640449187718332
step: 20, loss: 0.00019652635091915727
step: 30, loss: 0.006796615663915873
step: 40, loss: 0.027109965682029724
step: 50, loss: 0.004519160836935043
step: 60, loss: 0.002249865559861064
step: 70, loss: 0.0014444055268540978
step: 80, loss: 0.0007584225386381149
step: 90, loss: 0.035169217735528946
step: 100, loss: 0.0007470950949937105
step: 110, loss: 0.03280596435070038
step: 120, loss: 0.00442248210310936
step: 130, loss: 0.0007711885264143348
step: 140, loss: 0.016188351437449455
step: 150, loss: 0.0029032351449131966
step: 160, loss: 0.007439950481057167
step: 170, loss: 0.02723340317606926
step: 180, loss: 0.010354161262512207
step: 190, loss: 0.0006140315090306103
step: 200, loss: 0.005677712149918079
step: 210, loss: 0.0006159418262541294
step: 220, loss: 0.00028472847770899534
step: 230, loss: 0.0009777916129678488
step: 240, loss: 0.00304793706163764
step: 250, loss: 0.0002296697930432856
step: 260, loss: 0.0032281740568578243
step: 270, loss: 0.0067051914520561695
step: 280, loss: 0.05785346403717995
step: 290, loss: 0.00015644478844478726
step: 300, loss: 0.0012667668052017689
step: 310, loss: 0.00017496325017418712
```

```
step: 320, loss: 0.000680179800838232
step: 330, loss: 0.0006503754411824048
step: 340, loss: 0.010486113838851452
step: 350, loss: 0.08833753317594528
step: 360, loss: 0.09531933069229126
step: 370, loss: 0.2678813636302948
step: 380, loss: 0.07434917986392975
step: 390, loss: 0.019718097522854805
step: 400, loss: 0.011777306906878948
step: 410, loss: 0.0011363767553120852
step: 420, loss: 0.005427567753940821
step: 430, loss: 0.0007578323711641133
step: 440, loss: 0.002901021158322692
step: 450, loss: 0.0003026149934157729
step: 460, loss: 0.00028774573002010584
step: 470, loss: 0.024429403245449066
step: 480, loss: 0.0034734169021248817
step: 490, loss: 0.008164179511368275
step: 500, loss: 0.005986376665532589
step: 510, loss: 0.03896363079547882
step: 520, loss: 0.008076557889580727
step: 530, loss: 0.007098368369042873
/usr/local/lib/python3.7/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=0.9423165666820489, f1=0.943344081068632,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0043135653249919415
step: 10, loss: 0.0008622102905064821
step: 20, loss: 0.0003116173029411584
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 30, loss: 0.0003894123656209558
step: 40, loss: 0.007440793327987194
step: 50, loss: 0.00028682543779723346
step: 60, loss: 0.0010977658675983548
step: 70, loss: 0.018743183463811874
step: 80, loss: 0.0011106299934908748
step: 90, loss: 0.0006779220420867205
step: 100, loss: 0.0055571976117789745
step: 110, loss: 0.0037975243758410215
step: 120, loss: 0.00028121029026806355
step: 130, loss: 0.0017911494942381978
step: 140, loss: 0.0010685338638722897
step: 150, loss: 0.00014016011846251786
step: 160, loss: 0.0002342785446671769
```

```
step: 170, loss: 0.0003242537204641849
step: 180, loss: 0.0003446716582402587
step: 190, loss: 0.01667405664920807
step: 200, loss: 0.00019495545711833984
step: 210, loss: 0.0006848248303867877
step: 220, loss: 0.00044920292566530406
step: 230, loss: 0.000457068148534745
step: 240, loss: 0.00044978849473409355
step: 250, loss: 0.0002542174479458481
step: 260, loss: 0.014684912748634815
step: 270, loss: 0.0012034974060952663
step: 280, loss: 0.0004996774950996041
step: 290, loss: 0.00090873206499964
step: 300, loss: 0.0012322210241109133
step: 310, loss: 0.0029150526970624924
step: 320, loss: 0.08781635016202927
step: 330, loss: 0.0013959042262285948
step: 340, loss: 0.0006320042884908617
step: 350, loss: 0.035765837877988815
step: 360, loss: 0.0056761703453958035
step: 370, loss: 0.003345376579090953
step: 380, loss: 0.003772591706365347
step: 390, loss: 0.004477416165173054
step: 400, loss: 0.00014388348790816963
step: 410, loss: 0.04043320193886757
step: 420, loss: 0.12514281272888184
step: 430, loss: 0.0009654028690420091
step: 440, loss: 0.004724533762782812
step: 450, loss: 0.002403070917353034
step: 460, loss: 0.00969934742897749
step: 470, loss: 0.043263908475637436
step: 480, loss: 0.004885859787464142
step: 490, loss: 0.005026923026889563
step: 500, loss: 0.0037690557073801756
step: 510, loss: 0.020282192155718803
step: 520, loss: 0.0009906082414090633
step: 530, loss: 0.0015784921124577522
/usr/local/lib/python3.7/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=0.9407925407925408, f1=0.940570893776322,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001559103955514729
step: 10, loss: 0.002856605686247349
step: 20, loss: 0.006628059316426516
```

```
step: 30, loss: 0.0013031206326559186
step: 40, loss: 0.11580435186624527
step: 50, loss: 0.0063184103928506374
step: 60, loss: 0.008612020872533321
step: 70, loss: 0.008019482716917992
step: 80, loss: 0.07059615105390549
step: 90, loss: 0.0010204160353168845
step: 100, loss: 0.012752328999340534
step: 110, loss: 0.0007691952050663531
step: 120, loss: 0.0005900597316212952
step: 130, loss: 0.00018894441018346697
step: 140, loss: 0.009548638015985489
step: 150, loss: 0.007365132216364145
step: 160, loss: 0.0013036892050877213
step: 170, loss: 0.0023182244040071964
step: 180, loss: 0.00020836111798416823
step: 190, loss: 0.0007346761412918568
step: 200, loss: 0.00010788630606839433
step: 210, loss: 9.125072392635047e-05
step: 220, loss: 0.00018148173694498837
step: 230, loss: 0.0003111039404757321
step: 240, loss: 0.0033951285295188427
step: 250, loss: 0.0001914202148327604
step: 260, loss: 0.00011875748896272853
step: 270, loss: 0.0001726332411635667
step: 280, loss: 0.0009838148253038526
step: 290, loss: 0.00882949773222208
step: 300, loss: 0.008983880281448364
step: 310, loss: 0.0010360818123444915
step: 320, loss: 0.028609732165932655
step: 330, loss: 0.00020499664242379367
step: 340, loss: 0.025651009753346443
step: 350, loss: 0.00029679565341211855
step: 360, loss: 0.003465620568022132
step: 370, loss: 0.012308252975344658
step: 380, loss: 0.0018392184283584356
step: 390, loss: 0.00014658049622084945
step: 400, loss: 7.260608254000545e-05
step: 410, loss: 0.0018683437956497073
step: 420, loss: 8.145364699885249e-05
step: 430, loss: 0.0002018768573179841
step: 440, loss: 0.0003754459321498871
step: 450, loss: 9.173589933197945e-05
step: 460, loss: 0.0003591877466533333
step: 470, loss: 0.003740935120731592
step: 480, loss: 0.002218897920101881
step: 490, loss: 0.0017025393899530172
step: 500, loss: 0.00014029568410478532
```

```
step: 510, loss: 0.015203875489532948
step: 520, loss: 0.0001502472732681781
step: 530, loss: 4.8232581320917234e-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=0.9411764705882353, f1=0.9448244414044686,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009536569705232978
step: 10, loss: 0.0008744357037357986
step: 20, loss: 0.02262398600578308
step: 30, loss: 0.004320103675127029
step: 40, loss: 0.00010588644363451749
step: 50, loss: 0.00012170006812084466
step: 60, loss: 0.00016063862130977213
step: 70, loss: 4.420554978423752e-05
step: 80, loss: 4.50662337243557e-05
step: 90, loss: 0.00010859106987481937
step: 100, loss: 0.0011209186632186174
step: 110, loss: 0.0005400551599450409
step: 120, loss: 0.00013305858010426164
step: 130, loss: 0.00013900984777137637
step: 140, loss: 2.803200732159894e-05
step: 150, loss: 0.00029552902560681105
step: 160, loss: 0.00016690001939423382
step: 170, loss: 0.0002750442363321781
step: 180, loss: 4.894013545708731e-05
step: 190, loss: 6.160609336802736e-05
step: 200, loss: 0.00023492210311815143
step: 210, loss: 0.00020593423687387258
step: 220, loss: 0.0006275800988078117
step: 230, loss: 0.0001262175355805084
step: 240, loss: 5.562190926866606e-05
step: 250, loss: 0.00013303956075105816
step: 260, loss: 0.0015278563369065523
step: 270, loss: 4.5786404371028766e-05
step: 280, loss: 0.00013100057549308985
step: 290, loss: 0.00025175645714625716
step: 300, loss: 3.66403728548903e-05
step: 310, loss: 0.018556702882051468
step: 320, loss: 0.007686046417802572
step: 330, loss: 0.02638261392712593
step: 340, loss: 0.002366873202845454
step: 350, loss: 0.0019647672306746244
step: 360, loss: 0.00015546887880191207
```

```
step: 370, loss: 0.0004198838141746819
step: 380, loss: 0.00011660241580102593
step: 390, loss: 0.00043234601616859436
step: 400, loss: 0.06100337579846382
step: 410, loss: 0.002669458044692874
step: 420, loss: 0.00012900875299237669
step: 430, loss: 0.008523371070623398
step: 440, loss: 0.0023366690147668123
step: 450, loss: 0.000887086265720427
step: 460, loss: 0.0005695982254110277
step: 470, loss: 7.319681026274338e-05
step: 480, loss: 0.00019349371723365039
step: 490, loss: 8.762161451159045e-05
step: 500, loss: 3.670637306640856e-05
step: 510, loss: 3.260253288317472e-05
step: 520, loss: 0.005463343113660812
step: 530, loss: 0.00022903746867086738
/usr/local/lib/python3.7/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=0.9335810496980957, f1=0.9392675011590171,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.663496292778291e-05
step: 10, loss: 3.782381099881604e-05
step: 20, loss: 0.00016468315152451396
step: 30, loss: 0.0007869398104958236
step: 40, loss: 0.006261151749640703
step: 50, loss: 0.0010487496620044112
step: 60, loss: 0.0001787444925867021
step: 70, loss: 0.10129816830158234
step: 80, loss: 0.00020876432245131582
step: 90, loss: 0.007258119527250528
step: 100, loss: 0.00018895564426202327
step: 110, loss: 0.0007571967435069382
step: 120, loss: 0.00011714081483660266
step: 130, loss: 0.0024777199141681194
step: 140, loss: 0.00018149107927456498
step: 150, loss: 3.764588836929761e-05
step: 160, loss: 0.0008224945049732924
step: 170, loss: 0.0006925333291292191
step: 180, loss: 3.261332312831655e-05
step: 190, loss: 0.000248563737841323
step: 200, loss: 0.0032567400485277176
step: 210, loss: 0.00015139978495426476
step: 220, loss: 8.52751181810163e-05
```

```
step: 230, loss: 0.007076086010783911
step: 240, loss: 0.000691238499712199
step: 250, loss: 0.0004704639140982181
step: 260, loss: 0.0011779996566474438
step: 270, loss: 0.00038505185511894524
step: 280, loss: 4.205643563182093e-05
step: 290, loss: 9.08991860342212e-05
step: 300, loss: 0.000332859082845971
step: 310, loss: 3.1327643227996305e-05
step: 320, loss: 0.030133647844195366
step: 330, loss: 0.0024863029830157757
step: 340, loss: 0.10852473229169846
step: 350, loss: 0.0014611276565119624
step: 360, loss: 0.11214298754930496
step: 370, loss: 0.0024409655015915632
step: 380, loss: 0.00014436694618780166
step: 390, loss: 0.0030393165070563555
step: 400, loss: 0.0015035991091281176
step: 410, loss: 0.000788675679359585
step: 420, loss: 0.0004168566665612161
step: 430, loss: 0.00012090646487195045
step: 440, loss: 0.000167055957717821
step: 450, loss: 0.00011464247654657811
step: 460, loss: 5.34151477040723e-05
step: 470, loss: 8.278259338112548e-05
step: 480, loss: 0.0002614702971186489
step: 490, loss: 0.00023325972142629325
step: 500, loss: 0.009071051143109798
step: 510, loss: 0.0011503610294312239
step: 520, loss: 0.010595740750432014
step: 530, loss: 8.363146480405703e-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=0.9443166129774505, f1=0.9447236180904524,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.966905104694888e-05
step: 10, loss: 0.0005343835800886154
step: 20, loss: 0.0002455519570503384
step: 30, loss: 2.819535984599497e-05
step: 40, loss: 8.9154505985789e-05
step: 50, loss: 2.634393604239449e-05
step: 60, loss: 8.502772107021883e-05
step: 70, loss: 0.00010930465941783041
step: 80, loss: 0.0005982874426990747
```

```
step: 90, loss: 0.00015740716480650008
step: 100, loss: 0.00021801298134960234
step: 110, loss: 0.0006108001689426601
step: 120, loss: 0.0001303602912230417
step: 130, loss: 2.7483796657179482e-05
step: 140, loss: 5.639789014821872e-05
step: 150, loss: 4.742256351164542e-05
step: 160, loss: 0.0001757270802045241
step: 170, loss: 0.00015707364946138114
step: 180, loss: 0.015440180897712708
step: 190, loss: 0.00020605266036000103
step: 200, loss: 0.0028413960244506598
step: 210, loss: 6.156231393106282e-05
step: 220, loss: 4.699438431998715e-05
step: 230, loss: 5.815845725010149e-05
step: 240, loss: 3.256520358263515e-05
step: 250, loss: 0.00014721181651111692
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 260, loss: 0.0018625042866915464
step: 270, loss: 3.611339343478903e-05
step: 280, loss: 2.8031672627548687e-05
step: 290, loss: 4.5009706809651107e-05
step: 300, loss: 0.0016112817684188485
step: 310, loss: 7.613937486894429e-05
step: 320, loss: 0.00045107395271770656
step: 330, loss: 0.0006159748882055283
step: 340, loss: 3.6448807804845273e-05
step: 350, loss: 2.4310536900884472e-05
step: 360, loss: 6.999953620834276e-05
step: 370, loss: 0.000296143552986905
step: 380, loss: 8.635919948574156e-05
step: 390, loss: 2.8698585083475336e-05
step: 400, loss: 2.986786057590507e-05
step: 410, loss: 5.893449269933626e-05
step: 420, loss: 0.005221304018050432
step: 430, loss: 5.475064608617686e-05
step: 440, loss: 0.0004022255016025156
step: 450, loss: 0.00030145468190312386
step: 460, loss: 4.400309990160167e-05
step: 470, loss: 2.6385358069092035e-05
step: 480, loss: 5.2521918405545875e-05
step: 490, loss: 0.05701654404401779
step: 500, loss: 0.0014240503078326583
step: 510, loss: 0.0004014937730971724
step: 520, loss: 7.320637814700603e-05
step: 530, loss: 2.448874874971807e-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=0.943502824858757, f1=0.9460347254809948,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003855114337056875
step: 10, loss: 0.0021756698843091726
step: 20, loss: 1.684179369476624e-05
step: 30, loss: 0.00014431557792704552
step: 40, loss: 0.00010539919458096847
step: 50, loss: 0.001424479647539556
step: 60, loss: 0.004042464774101973
step: 70, loss: 2.901849256886635e-05
step: 80, loss: 8.312272984767333e-05
step: 90, loss: 0.00010055279562948272
step: 100, loss: 5.721927300328389e-05
step: 110, loss: 0.00033953526872210205
step: 120, loss: 0.0002540967834647745
step: 130, loss: 3.7436115235323086e-05
step: 140, loss: 3.507140718284063e-05
step: 150, loss: 2.5055600417545065e-05
step: 160, loss: 4.775191337103024e-05
step: 170, loss: 8.12650760053657e-05
step: 180, loss: 2.85902733594412e-05
step: 190, loss: 0.0002546315372455865
step: 200, loss: 0.0018620737828314304
step: 210, loss: 2.374822179262992e-05
step: 220, loss: 5.5080025049392134e-05
step: 230, loss: 2.8581092919921502e-05
step: 240, loss: 3.276141796959564e-05
step: 250, loss: 3.2768566597951576e-05
step: 260, loss: 3.109641693299636e-05
step: 270, loss: 0.0011845570988953114
step: 280, loss: 0.00026852902374230325
step: 290, loss: 0.0018261607037857175
step: 300, loss: 0.00041394177242182195
step: 310, loss: 3.9189719245769083e-05
step: 320, loss: 0.013106059283018112
step: 330, loss: 0.0005099182017147541
step: 340, loss: 4.231594721204601e-05
step: 350, loss: 0.0008183716563507915
step: 360, loss: 5.839366349391639e-05
step: 370, loss: 1.7437831047573127e-05
step: 380, loss: 0.00056530034635216
step: 390, loss: 0.09391002357006073
step: 400, loss: 1.716218321234919e-05
step: 410, loss: 0.0008115450036711991
```

```
step: 420, loss: 0.00022043430362828076
step: 430, loss: 0.000277270853985101
step: 440, loss: 3.8886348193045706e-05
step: 450, loss: 3.691926394822076e-05
step: 460, loss: 0.0025189025327563286
step: 470, loss: 4.125517079955898e-05
step: 480, loss: 2.4180357286240906e-05
step: 490, loss: 0.0034808970522135496
step: 500, loss: 1.9397279174881987e-05
step: 510, loss: 1.91365325008519e-05
step: 520, loss: 1.9915019947802648e-05
step: 530, loss: 2.7334832338965498e-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=0.9440820130475303, f1=0.9458583988894032,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014341590576805174
step: 10, loss: 3.428082709433511e-05
step: 20, loss: 2.8086773454560898e-05
step: 30, loss: 1.9125383914797567e-05
step: 40, loss: 0.00013304046296980232
step: 50, loss: 0.0005999717977829278
step: 60, loss: 2.7725924155674875e-05
step: 70, loss: 4.501644070842303e-05
step: 80, loss: 0.002464944962412119
step: 90, loss: 3.463837128947489e-05
step: 100, loss: 8.142968727042899e-05
step: 110, loss: 1.7001984815578908e-05
step: 120, loss: 0.009718780405819416
step: 130, loss: 3.0407833037315868e-05
step: 140, loss: 4.645062654162757e-05
step: 150, loss: 0.0002372638409724459
step: 160, loss: 0.0002526401658542454
step: 170, loss: 1.877514660009183e-05
step: 180, loss: 3.256846684962511e-05
step: 190, loss: 5.689175668521784e-05
step: 200, loss: 3.0873867217451334e-05
step: 210, loss: 0.000251517427386716
step: 220, loss: 1.9408480511629023e-05
step: 230, loss: 0.0017172860680148005
step: 240, loss: 2.4175862563424744e-05
step: 250, loss: 2.6120813345187344e-05
step: 260, loss: 3.765614746953361e-05
step: 270, loss: 3.6431480111787096e-05
```

```
step: 280, loss: 4.6941800974309444e-05
step: 290, loss: 0.0028407725039869547
step: 300, loss: 5.2264407713664696e-05
step: 310, loss: 0.00020042041433043778
step: 320, loss: 3.207197005394846e-05
step: 330, loss: 2.342040534131229e-05
step: 340, loss: 2.5506164092803374e-05
step: 350, loss: 0.03420061990618706
step: 360, loss: 0.0004867423267569393
step: 370, loss: 2.03508316189982e-05
step: 380, loss: 2.6363031793152913e-05
step: 390, loss: 3.135786755592562e-05
step: 400, loss: 2.6087373043992557e-05
step: 410, loss: 0.00018424783775117248
step: 420, loss: 0.0006673047901131213
step: 430, loss: 9.536392462905496e-05
step: 440, loss: 7.205624569905922e-05
step: 450, loss: 0.00012442408478818834
step: 460, loss: 0.00023910420713946223
step: 470, loss: 0.00011845496919704601
step: 480, loss: 0.00020949705503880978
step: 490, loss: 3.0484656235785224e-05
step: 500, loss: 0.004091642796993256
step: 510, loss: 2.74247613560874e-05
step: 520, loss: 0.017065098509192467
step: 530, loss: 0.0003335805085953325
/usr/local/lib/python3.7/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=0.9446254071661238, f1=0.9451360073766714,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011275694705545902
step: 10, loss: 2.1318855942809023e-05
step: 20, loss: 0.03206916153430939
step: 30, loss: 9.832559589995071e-05
step: 40, loss: 0.00035872991429641843
step: 50, loss: 5.849422450410202e-05
step: 60, loss: 7.512085721828043e-05
step: 70, loss: 3.640647264546715e-05
step: 80, loss: 2.2026992155588232e-05
step: 90, loss: 0.00011604364408412948
step: 100, loss: 1.6435689758509398e-05
step: 110, loss: 0.00024343874247279018
step: 120, loss: 3.8818401662865654e-05
step: 130, loss: 0.002219222253188491
```

```
step: 140, loss: 7.933467713883147e-05
step: 150, loss: 3.851636211038567e-05
step: 160, loss: 2.4016118914005347e-05
step: 170, loss: 0.000526159128639847
step: 180, loss: 3.8980069803074e-05
step: 190, loss: 9.558539022691548e-05
step: 200, loss: 0.001671376870945096
step: 210, loss: 0.000993021298199892
step: 220, loss: 1.6566122212680057e-05
step: 230, loss: 0.00037728165625594556
step: 240, loss: 4.1967094148276374e-05
step: 250, loss: 0.031527403742074966
step: 260, loss: 0.0005341320647858083
step: 270, loss: 0.005277745891362429
step: 280, loss: 0.00019648554734885693
step: 290, loss: 0.00010278508125338703
step: 300, loss: 0.000307450070977211
step: 310, loss: 7.17382354196161e-05
step: 320, loss: 7.706035830779001e-05
step: 330, loss: 0.00026752945268526673
step: 340, loss: 0.00022148100833874196
step: 350, loss: 0.01981191523373127
step: 360, loss: 1.5381421690108255e-05
step: 370, loss: 1.99556761799613e-05
step: 380, loss: 0.00041843278449960053
step: 390, loss: 0.0013293868396431208
step: 400, loss: 1.6148900613188744e-05
step: 410, loss: 5.574262831942178e-05
step: 420, loss: 2.022030639636796e-05
step: 430, loss: 0.005097547080367804
step: 440, loss: 5.6646978919161484e-05
step: 450, loss: 0.0003632009029388428
step: 460, loss: 1.706519651634153e-05
step: 470, loss: 0.015055302530527115
step: 480, loss: 0.0010518906638026237
step: 490, loss: 6.788522296119481e-05
step: 500, loss: 0.00016223454440478235
step: 510, loss: 1.943815550475847e-05
step: 520, loss: 1.621965566300787e-05
step: 530, loss: 2.045520523097366e-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=0.9464450600184672,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 3.425059185246937e-05
step: 10, loss: 6.559918256243691e-05
step: 20, loss: 1.3463048162520863e-05
step: 30, loss: 4.0954750147648156e-05
step: 40, loss: 0.0002710747648961842
step: 50, loss: 0.00036199306487105787
step: 60, loss: 2.260068322357256e-05
step: 70, loss: 1.6305348253808916e-05
step: 80, loss: 0.0009246374829672277
step: 90, loss: 1.2189031622256152e-05
step: 100, loss: 3.990990808233619e-05
step: 110, loss: 1.978816362679936e-05
step: 120, loss: 0.0002331729920115322
step: 130, loss: 4.125963096157648e-05
step: 140, loss: 0.0008008677395991981
step: 150, loss: 0.007252611219882965
step: 160, loss: 0.0007442852947860956
step: 170, loss: 0.00023906896240077913
step: 180, loss: 0.016385367140173912
step: 190, loss: 4.514468309935182e-05
step: 200, loss: 9.759952808963135e-05
step: 210, loss: 0.00019615681958384812
step: 220, loss: 2.4108972866088152e-05
step: 230, loss: 0.001908954232931137
step: 240, loss: 5.2359115215949714e-05
step: 250, loss: 2.7610944016487338e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 260, loss: 1.2468439308577217e-05
step: 270, loss: 0.0032698907889425755
step: 280, loss: 2.1136456780368462e-05
step: 290, loss: 0.00020841050718444533
step: 300, loss: 0.00010425265645608306
step: 310, loss: 4.798379814019427e-05
step: 320, loss: 0.00016088168194983155
step: 330, loss: 0.00020347017562016845
step: 340, loss: 3.2441846997244284e-05
step: 350, loss: 0.00017713781562633812
step: 360, loss: 0.0027247874531894922
step: 370, loss: 3.764717257581651e-05
step: 380, loss: 2.843573565769475e-05
step: 390, loss: 4.359638478490524e-05
step: 400, loss: 4.268623888492584e-05
step: 410, loss: 0.0005875737406313419
step: 420, loss: 1.5206461284833495e-05
step: 430, loss: 6.233371095731854e-05
step: 440, loss: 1.4122395441518165e-05
step: 450, loss: 4.75617780466564e-05
step: 460, loss: 0.001568888546898961
```

```
step: 470, loss: 1.762020656315144e-05
step: 480, loss: 1.4334746992972214e-05
step: 490, loss: 3.348755490151234e-05
step: 500, loss: 0.0001866774255177006
step: 510, loss: 2.502510506019462e-05
step: 520, loss: 3.3963755413424224e-05
step: 530, loss: 4.4072385207982734e-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=0.944157672454247, f1=0.9434315100514259,
best_f1=0.9422632794457274
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 1.5411284039146267e-05
step: 10, loss: 2.93017401418183e-05
step: 20, loss: 2.8488062525866553e-05
step: 30, loss: 0.00017534567456063886
step: 40, loss: 1.8383598217042163e-05
step: 50, loss: 0.0006747012957930565
step: 60, loss: 2.7611993573373184e-05
step: 70, loss: 0.00029118676320649683
step: 80, loss: 1.9121318473480642e-05
step: 90, loss: 1.8432148863212205e-05
step: 100, loss: 2.2600306692766026e-05
step: 110, loss: 0.013585918582975864
step: 120, loss: 0.0018488230416551232
step: 130, loss: 0.00017102682613767684
step: 140, loss: 1.3097974260745104e-05
step: 150, loss: 6.079225931898691e-05
step: 160, loss: 0.00020141353888902813
step: 170, loss: 0.00015040051948744804
step: 180, loss: 0.00011862207611557096
step: 190, loss: 5.308729305397719e-05
step: 200, loss: 2.9244094548630528e-05
step: 210, loss: 3.9836821088101715e-05
step: 220, loss: 1.080324909707997e-05
step: 230, loss: 1.913266532937996e-05
step: 240, loss: 1.6204798157559708e-05
step: 250, loss: 4.844671639148146e-05
step: 260, loss: 2.8042622943758033e-05
step: 270, loss: 0.00010285676398780197
step: 280, loss: 0.00036790952435694635
step: 290, loss: 0.0017209105426445603
step: 300, loss: 0.004291472490876913
step: 310, loss: 0.0003981771878898144
step: 320, loss: 2.590831536508631e-05
```

```
step: 330, loss: 8.512799104209989e-05
step: 340, loss: 0.00014466853463090956
step: 350, loss: 2.077916360576637e-05
step: 360, loss: 2.0879686417174526e-05
step: 370, loss: 1.6528776541235857e-05
step: 380, loss: 2.652186776685994e-05
step: 390, loss: 1.5195200830930844e-05
step: 400, loss: 1.775422424543649e-05
step: 410, loss: 0.05143557861447334
step: 420, loss: 3.2479434594279155e-05
step: 430, loss: 2.57701103691943e-05
step: 440, loss: 5.523276195162907e-05
step: 450, loss: 0.00032373497379012406
step: 460, loss: 1.4282576557889115e-05
step: 470, loss: 0.00020677810243796557
step: 480, loss: 0.00019700730626937002
step: 490, loss: 0.0004601088003255427
step: 500, loss: 0.036081381142139435
step: 510, loss: 2.190742270613555e-05
step: 520, loss: 2.7993031835649163e-05
step: 530, loss: 2.5770237698452547e-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=0.9439555349698935, f1=0.9480459770114943,
best_f1=0.9422632794457274
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
```

```
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:13, 414.58it/s]
load_f1 = 0.9442653155228006
real_f1 = 0.9459211732355638
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 437.83it/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: 0.8283816576004028
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06451018154621124
step: 20, loss: 0.3736538290977478
step: 30, loss: 0.3598952293395996
step: 40, loss: 0.4677610695362091
step: 50, loss: 0.27872270345687866
step: 60, loss: 0.3659786581993103
step: 70, loss: 0.18104061484336853
step: 80, loss: 0.3111690282821655
step: 90, loss: 0.4001881778240204
step: 100, loss: 0.10351169109344482
```

```
step: 110, loss: 0.38091611862182617
step: 120, loss: 0.242440864443779
step: 130, loss: 0.17936378717422485
step: 140, loss: 0.18329420685768127
step: 150, loss: 0.24270080029964447
step: 160, loss: 0.20978012681007385
step: 170, loss: 0.1799197942018509
step: 180, loss: 0.14978095889091492
step: 190, loss: 0.13876554369926453
step: 200, loss: 0.13346783816814423
step: 210, loss: 0.3673274517059326
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6245353159851301, f1=0.6731898238747555,
best_f1=0.6731898238747555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08247844874858856
step: 10, loss: 0.07171818614006042
step: 20, loss: 0.1430307775735855
step: 30, loss: 0.16963739693164825
step: 40, loss: 0.0927463173866272
step: 50, loss: 0.170502707362175
step: 60, loss: 0.05887678265571594
step: 70, loss: 0.16720569133758545
step: 80, loss: 0.13538312911987305
step: 90, loss: 0.09333410859107971
step: 100, loss: 0.06433691084384918
step: 110, loss: 0.09157067537307739
step: 120, loss: 0.17362487316131592
step: 130, loss: 0.21221141517162323
step: 140, loss: 0.16438505053520203
step: 150, loss: 0.1419035792350769
step: 160, loss: 0.10781551897525787
step: 170, loss: 0.20252811908721924
step: 180, loss: 0.24937774240970612
step: 190, loss: 0.0784483477473259
step: 200, loss: 0.1458282470703125
step: 210, loss: 0.1650238037109375
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6833013435700576, f1=0.7129094412331405,
best_f1=0.7129094412331405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13909800350666046
step: 10, loss: 0.20046892762184143
step: 20, loss: 0.24984446167945862
step: 30, loss: 0.04725629836320877
step: 40, loss: 0.15333044528961182
step: 50, loss: 0.16535986959934235
step: 60, loss: 0.21237210929393768
step: 70, loss: 0.10723382234573364
step: 80, loss: 0.04759861156344414
step: 90, loss: 0.13585810363292694
step: 100, loss: 0.045494407415390015
step: 110, loss: 0.06632021814584732
step: 120, loss: 0.18465542793273926
step: 130, loss: 0.1084442138671875
step: 140, loss: 0.25442248582839966
step: 150, loss: 0.0920647382736206
step: 160, loss: 0.09393741935491562
step: 170, loss: 0.07925301045179367
step: 180, loss: 0.058212365955114365
step: 190, loss: 0.14330895245075226
step: 200, loss: 0.10633990168571472
step: 210, loss: 0.1424049735069275
/usr/local/lib/python3.7/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=0.6797385620915033, f1=0.721868365180467,
best_f1=0.7129094412331405
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0976492390036583
step: 10, loss: 0.021683737635612488
step: 20, loss: 0.09251256287097931
step: 30, loss: 0.12274280190467834
step: 40, loss: 0.028496889397501945
step: 50, loss: 0.09849689155817032
step: 60, loss: 0.041475869715213776
step: 70, loss: 0.13477183878421783
step: 80, loss: 0.10710785537958145
```

```
step: 90, loss: 0.012661324813961983
step: 100, loss: 0.08301536738872528
step: 110, loss: 0.10531695932149887
step: 120, loss: 0.012250645086169243
step: 130, loss: 0.06369901448488235
step: 140, loss: 0.23180685937404633
step: 150, loss: 0.11063247919082642
step: 160, loss: 0.16229534149169922
step: 170, loss: 0.03244921565055847
step: 180, loss: 0.18037289381027222
step: 190, loss: 0.05013453587889671
step: 200, loss: 0.04448411241173744
step: 210, loss: 0.01169948372989893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6935483870967742, f1=0.6963562753036439,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03286373242735863
step: 10, loss: 0.08674313127994537
step: 20, loss: 0.0063585396856069565
step: 30, loss: 0.04693640023469925
step: 40, loss: 0.21348197758197784
step: 50, loss: 0.09123124182224274
step: 60, loss: 0.1163933202624321
step: 70, loss: 0.05246863514184952
step: 80, loss: 0.017567893490195274
step: 90, loss: 0.029298385605216026
step: 100, loss: 0.02232341468334198
step: 110, loss: 0.011847324669361115
step: 120, loss: 0.006853078026324511
step: 130, loss: 0.016335060819983482
step: 140, loss: 0.014692436903715134
step: 150, loss: 0.0420108363032341
step: 160, loss: 0.014566768892109394
step: 170, loss: 0.024561231955885887
step: 180, loss: 0.20020005106925964
step: 190, loss: 0.1488175243139267
step: 200, loss: 0.07681875675916672
step: 210, loss: 0.01785554178059101
/usr/local/lib/python3.7/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=0.6789366053169735, f1=0.6858316221765914,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08821888267993927
step: 10, loss: 0.06445865333080292
step: 20, loss: 0.006743620615452528
step: 30, loss: 0.04710251837968826
step: 40, loss: 0.00255612563341856
step: 50, loss: 0.03249150142073631
step: 60, loss: 0.09171182662248611
step: 70, loss: 0.0035194975789636374
step: 80, loss: 0.09168877452611923
step: 90, loss: 0.035081058740615845
step: 100, loss: 0.018317757174372673
step: 110, loss: 0.023703187704086304
step: 120, loss: 0.04663701355457306
step: 130, loss: 0.00764313293620944
step: 140, loss: 0.04407806321978569
step: 150, loss: 0.006389403250068426
step: 160, loss: 0.053323954343795776
step: 170, loss: 0.08629120886325836
step: 180, loss: 0.00315662263892591
step: 190, loss: 0.11552207171916962
step: 200, loss: 0.007491816766560078
step: 210, loss: 0.040137361735105515
/usr/local/lib/python3.7/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=0.6736401673640167, f1=0.6889352818371608,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015951447188854218
step: 10, loss: 0.014759324491024017
step: 20, loss: 0.01622452400624752
step: 30, loss: 0.042186591774225235
step: 40, loss: 0.08333863317966461
step: 50, loss: 0.056638944894075394
step: 60, loss: 0.06766320765018463
step: 70, loss: 0.002883684355765581
step: 80, loss: 0.006877687759697437
step: 90, loss: 0.011590678244829178
step: 100, loss: 0.0022054894361644983
```

```
step: 110, loss: 0.05470883473753929
step: 120, loss: 0.03492588549852371
step: 130, loss: 0.0005503717693500221
step: 140, loss: 0.008963057771325111
step: 150, loss: 0.11512116342782974
step: 160, loss: 0.008641294203698635
step: 170, loss: 0.01934673823416233
step: 180, loss: 0.0016285396413877606
step: 190, loss: 0.000718633527867496
step: 200, loss: 0.0033664554357528687
step: 210, loss: 0.029870085418224335
/usr/local/lib/python3.7/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=0.683111954459203, f1=0.6972477064220184,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03449571505188942
step: 10, loss: 0.031248468905687332
step: 20, loss: 0.028214003890752792
step: 30, loss: 0.042143769562244415
step: 40, loss: 0.030218634754419327
step: 50, loss: 0.018799349665641785
step: 60, loss: 0.129940003156662
step: 70, loss: 0.0026006100233644247
step: 80, loss: 0.014726969413459301
step: 90, loss: 0.0038732148241251707
step: 100, loss: 0.001354214851744473
step: 110, loss: 0.0009687747806310654
step: 120, loss: 0.0011685675708577037
step: 130, loss: 0.08702509850263596
step: 140, loss: 0.0037646570708602667
step: 150, loss: 0.06056710332632065
step: 160, loss: 0.03604729473590851
step: 170, loss: 0.015358097851276398
step: 180, loss: 0.0984613224864006
step: 190, loss: 0.043040681630373
step: 200, loss: 0.009531158022582531
step: 210, loss: 0.07686732709407806
/usr/local/lib/python3.7/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=0.6719681908548708, f1=0.7081712062256809,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01322835311293602
step: 10, loss: 0.023470543324947357
step: 20, loss: 0.019195282831788063
step: 30, loss: 0.02046823874115944
step: 40, loss: 0.18302294611930847
step: 50, loss: 0.001248826622031629
step: 60, loss: 0.0026386934333254957
step: 70, loss: 0.0007744574104435742
step: 80, loss: 0.000780714035499841
step: 90, loss: 0.13491317629814148
step: 100, loss: 0.003174275392666459
step: 110, loss: 0.01845393516123295
step: 120, loss: 0.0025595747865736485
step: 130, loss: 0.0003249887377023697
step: 140, loss: 0.0122957369312644
step: 150, loss: 0.003037916263565421
step: 160, loss: 0.010059352032840252
step: 170, loss: 0.012581588700413704
step: 180, loss: 0.03506019711494446
step: 190, loss: 0.058623477816581726
step: 200, loss: 0.021230798214673996
step: 210, loss: 0.25506913661956787
/usr/local/lib/python3.7/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=0.6771653543307088, f1=0.7117988394584138,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017911366885527968
step: 10, loss: 0.003008611500263214
step: 20, loss: 0.011646505445241928
step: 30, loss: 0.005432570353150368
step: 40, loss: 0.03141488879919052
step: 50, loss: 8.41337678139098e-05
step: 60, loss: 0.00013156670320313424
step: 70, loss: 0.006182897835969925
step: 80, loss: 6.339101673802361e-05
step: 90, loss: 0.01238197647035122
step: 100, loss: 0.00030031686765141785
step: 110, loss: 0.003498926991596818
step: 120, loss: 0.006717722862958908
step: 130, loss: 0.0023248197976499796
step: 140, loss: 0.00028718909015879035
step: 150, loss: 0.021961243823170662
step: 160, loss: 0.031198665499687195
```

```
step: 170, loss: 0.014839250594377518
step: 180, loss: 0.0005177645944058895
step: 190, loss: 0.06934940069913864
step: 200, loss: 0.030254870653152466
step: 210, loss: 0.025595203042030334
/usr/local/lib/python3.7/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=0.6890756302521008, f1=0.7063655030800821,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008063104934990406
step: 10, loss: 0.02772274985909462
step: 20, loss: 0.00015240287757478654
step: 30, loss: 0.0406690314412117
step: 40, loss: 0.03213471546769142
step: 50, loss: 0.004685480147600174
step: 60, loss: 0.0009651752770878375
step: 70, loss: 0.000535259663593024
step: 80, loss: 0.026000462472438812
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.0005489249597303569
step: 100, loss: 0.010054241865873337
step: 110, loss: 0.0096476711332798
step: 120, loss: 0.00024208294053096324
step: 130, loss: 0.03277161717414856
step: 140, loss: 0.08480895310640335
step: 150, loss: 0.035167671740055084
step: 160, loss: 0.00013690273044630885
step: 170, loss: 0.07366897165775299
step: 180, loss: 0.0040075726959854364
step: 190, loss: 0.011523803696036339
step: 200, loss: 0.0004770660598296672
step: 210, loss: 0.002775314263999462
/usr/local/lib/python3.7/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=0.6910569105691057, f1=0.6996047430830039,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002929984766524285
step: 10, loss: 0.0005120572168380022
step: 20, loss: 0.00016429867537226528
step: 30, loss: 0.003433129983022809
```

```
step: 40, loss: 0.00030598975718021393
step: 50, loss: 0.03884045034646988
step: 60, loss: 0.0002040254621533677
step: 70, loss: 0.1271730363368988
step: 80, loss: 0.0027719212230294943
step: 90, loss: 0.00013019393372815102
step: 100, loss: 0.00012405418965499848
step: 110, loss: 0.0004915056051686406
step: 120, loss: 0.0004158442316111177
step: 130, loss: 0.002716486807912588
step: 140, loss: 0.0007100393995642662
step: 150, loss: 0.00021114861010573804
step: 160, loss: 0.0003607221005950123
step: 170, loss: 0.013432436622679234
step: 180, loss: 0.00545845041051507
step: 190, loss: 0.04472961649298668
step: 200, loss: 0.08239539712667465
step: 210, loss: 0.08571892976760864
/usr/local/lib/python3.7/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=0.673728813559322, f1=0.6931106471816284,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001308140199398622
step: 10, loss: 0.00018394114158581942
step: 20, loss: 0.00022009597159922123
step: 30, loss: 0.006703246384859085
step: 40, loss: 0.00011308079410810024
step: 50, loss: 0.00010471084533492103
step: 60, loss: 0.006843185517936945
step: 70, loss: 0.04588260129094124
step: 80, loss: 0.027141019701957703
step: 90, loss: 0.018105272203683853
step: 100, loss: 0.04774341732263565
step: 110, loss: 0.001391323865391314
step: 120, loss: 0.007606359198689461
step: 130, loss: 0.0016953928861767054
step: 140, loss: 0.01422026939690113
step: 150, loss: 0.00011154141975566745
step: 160, loss: 0.01630130596458912
step: 170, loss: 0.02331126295030117
step: 180, loss: 0.004099222365766764
step: 190, loss: 7.057914626784623e-05
step: 200, loss: 0.0007186783477663994
step: 210, loss: 0.00016092555597424507
```

```
/usr/local/lib/python3.7/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=0.676, f1=0.7019607843137254, best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01685858704149723
step: 10, loss: 0.00011531810741871595
step: 20, loss: 0.00010072255827253684
step: 30, loss: 0.00014188805653247982
step: 40, loss: 0.001465512439608574
step: 50, loss: 0.008141026832163334
step: 60, loss: 0.008079078048467636
step: 70, loss: 0.0005643764743581414
step: 80, loss: 0.09222810715436935
step: 90, loss: 0.0003942280600313097
step: 100, loss: 0.00045387790305539966
step: 110, loss: 0.0020758402533829212
step: 120, loss: 8.312442514579743e-05
step: 130, loss: 6.529658276122063e-05
step: 140, loss: 7.396484579658136e-05
step: 150, loss: 0.0002448005834594369
step: 160, loss: 0.00047158473171293736
step: 170, loss: 8.997685654321685e-05
step: 180, loss: 0.00595104182139039
step: 190, loss: 0.09421281516551971
step: 200, loss: 7.861160702304915e-05
step: 210, loss: 0.022007256746292114
/usr/local/lib/python3.7/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=0.6567164179104478, f1=0.7014613778705637,
best_f1=0.6963562753036439
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04405180364847183
step: 10, loss: 0.006777493748813868
step: 20, loss: 8.105648157652467e-05
step: 30, loss: 0.00035011779982596636
step: 40, loss: 0.0007815298158675432
step: 50, loss: 9.521700849290937e-05
step: 60, loss: 0.0008891801699064672
step: 70, loss: 0.000252630707109347
step: 80, loss: 0.00017558789113536477
step: 90, loss: 0.00021960632875561714
step: 100, loss: 8.468082523904741e-05
```

```
step: 110, loss: 0.0002632474934216589
step: 120, loss: 0.00013765112089458853
step: 130, loss: 0.004312894772738218
step: 140, loss: 0.00016996274644043297
step: 150, loss: 0.0002045824658125639
step: 160, loss: 0.0005094575462862849
step: 170, loss: 0.00948238279670477
step: 180, loss: 0.03013727255165577
step: 190, loss: 0.0010037592146545649
step: 200, loss: 0.013959511183202267
step: 210, loss: 0.006187762133777142
/usr/local/lib/python3.7/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=0.6596194503171248, f1=0.709278350515464,
best_f1=0.6963562753036439
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.")
2293it [00:02, 777.59it/s]
load_f1 = 0.6904761904761906
```

```
real_f1 = 0.6935483870967742
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 436.81it/s]
```

2.1.5 Walmart-Amazon - Running the matcher

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

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

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

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

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
```

```
loss_scale               : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                  : True
opt_level                : O2
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`.  Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
  "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8578106760978699
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1703292429447174
step: 20, loss: 0.15149548649787903
step: 30, loss: 0.5019934773445129
step: 40, loss: 0.2473909854888916
step: 50, loss: 0.30716490745544434
step: 60, loss: 0.3572348654270172
step: 70, loss: 0.1734103411436081
step: 80, loss: 0.4970797300338745
step: 90, loss: 0.2247702032327652
step: 100, loss: 0.2150837779045105
step: 110, loss: 0.24344435334205627
step: 120, loss: 0.4272181987762451
step: 130, loss: 0.34468021988868713
step: 140, loss: 0.3254714012145996
step: 150, loss: 0.26231279969215393
step: 160, loss: 0.20817461609840393
step: 170, loss: 0.3692666292190552
step: 180, loss: 0.3549151122570038
step: 190, loss: 0.13735459744930267
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4528301886792453, f1=0.4647887323943662,
best_f1=0.4647887323943662
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29308319091796875
step: 10, loss: 0.03247221186757088
step: 20, loss: 0.08833735436201096
step: 30, loss: 0.22728317975997925
step: 40, loss: 0.5098658204078674
step: 50, loss: 0.2789206802845001
step: 60, loss: 0.16633902490139008
step: 70, loss: 0.1780014932155609
step: 80, loss: 0.13395994901657104
step: 90, loss: 0.07589641958475113
step: 100, loss: 0.2776201665401459
step: 110, loss: 0.1332039088010788
step: 120, loss: 0.20706906914710999
step: 130, loss: 0.12316713482141495
step: 140, loss: 0.23601336777210236
step: 150, loss: 0.010161756537854671
step: 160, loss: 0.12567901611328125
step: 170, loss: 0.2193821668624878
step: 180, loss: 0.2404487431049347
step: 190, loss: 0.1560758501291275
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6963788300835654, f1=0.7146814404432134,
best_f1=0.7146814404432134
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14496493339538574
step: 10, loss: 0.28660327196121216
step: 20, loss: 0.04573716223239899
step: 30, loss: 0.06811656802892685
step: 40, loss: 0.16150017082691193

```
step: 50, loss: 0.29698431491851807
step: 60, loss: 0.08306293189525604
step: 70, loss: 0.10597789287567139
step: 80, loss: 0.1898277848958969
step: 90, loss: 0.112861767411232
step: 100, loss: 0.10359810292720795
step: 110, loss: 0.17002302408218384
step: 120, loss: 0.02343583293259144
step: 130, loss: 0.05619258061051369
step: 140, loss: 0.09780562669038773
step: 150, loss: 0.05438046157360077
step: 160, loss: 0.1992301493883133
step: 170, loss: 0.07720549404621124
step: 180, loss: 0.0490674264729023
step: 190, loss: 0.13851182162761688
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7131367292225201, f1=0.7642276422764227,
best_f1=0.7642276422764227
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11257960647344589
step: 10, loss: 0.03980978950858116
step: 20, loss: 0.022445544600486755
step: 30, loss: 0.02552138641476631
step: 40, loss: 0.00309379701150262117
step: 50, loss: 0.0709901675581932
step: 60, loss: 0.11738673597574234
step: 70, loss: 0.022039752453565598
step: 80, loss: 0.047084975987672806
step: 90, loss: 0.008524822071194649
step: 100, loss: 0.0186778511852026
step: 110, loss: 0.021111659705638885
step: 120, loss: 0.07028260082006454
step: 130, loss: 0.2308485358953476
step: 140, loss: 0.03641153872013092
step: 150, loss: 0.025843186303973198
step: 160, loss: 0.028914345428347588
step: 170, loss: 0.04036618024110794
step: 180, loss: 0.13300158083438873
step: 190, loss: 0.04903046414256096
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7295285359801489, f1=0.7688442211055277,
best_f1=0.7688442211055277
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.041266415268182755
step: 10, loss: 0.015483198687434196
step: 20, loss: 0.11440150439739227
step: 30, loss: 0.006030069664120674
step: 40, loss: 0.104388028383255
step: 50, loss: 0.05101775750517845
step: 60, loss: 0.04363224655389786
step: 70, loss: 0.002129934262484312
step: 80, loss: 0.005336661357432604
step: 90, loss: 0.005770036950707436
step: 100, loss: 0.06899666041135788
step: 110, loss: 0.005738506093621254
step: 120, loss: 0.0006733571062795818
step: 130, loss: 0.018859008327126503
step: 140, loss: 0.02289537712931633
step: 150, loss: 0.019566697999835014
step: 160, loss: 0.008252182975411415
step: 170, loss: 0.006000129971653223
step: 180, loss: 0.038860660046339035
step: 190, loss: 0.14079006016254425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7421052631578948, f1=0.7679558011049724,
best_f1=0.7679558011049724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011778358370065689
step: 10, loss: 0.00038703574682585895
step: 20, loss: 0.0073320274241268635
step: 30, loss: 0.00249659875407815
step: 40, loss: 0.04620563983917236

```
step: 50, loss: 0.011367044411599636
step: 60, loss: 0.0005249017267487943
step: 70, loss: 0.00378695921972394
step: 80, loss: 0.018836649134755135
step: 90, loss: 0.003859440330415964
step: 100, loss: 0.0013934110756963491
step: 110, loss: 0.0016667862655594945
step: 120, loss: 0.02824046090245247
step: 130, loss: 0.0006975420983508229
step: 140, loss: 0.005014072638005018
step: 150, loss: 0.015116747468709946
step: 160, loss: 0.005761334206908941
step: 170, loss: 0.01433897577226162
step: 180, loss: 0.012122022919356823
step: 190, loss: 0.01213093288242817
/usr/local/lib/python3.7/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=0.73, f1=0.7571801566579635, best_f1=0.7679558011049724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003083935473114252
step: 10, loss: 0.005171111784875393
step: 20, loss: 0.0010476430179551244
step: 30, loss: 0.002210944192484021
step: 40, loss: 0.060767773538827896
step: 50, loss: 0.0399148166179657
step: 60, loss: 0.0045861429534852505
step: 70, loss: 0.0007644625729881227
step: 80, loss: 0.009744161739945412
step: 90, loss: 0.00433204835280776
step: 100, loss: 0.0030076089315116405
step: 110, loss: 0.001507198205217719
step: 120, loss: 0.0026463675312697887
step: 130, loss: 0.002750998130068183
step: 140, loss: 0.00035380147164687514
step: 150, loss: 0.0018951810197904706
step: 160, loss: 0.0038667921908199787
step: 170, loss: 0.0038275248371064663
step: 180, loss: 0.0013877777382731438
step: 190, loss: 0.00430708009749651
/usr/local/lib/python3.7/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=0.7354497354497355, f1=0.7439353099730459,
best_f1=0.7679558011049724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010137503268197179
step: 10, loss: 0.10000265389680862
step: 20, loss: 0.0005126229370944202
step: 30, loss: 0.0059363688342273235
step: 40, loss: 0.02131308615207672
step: 50, loss: 0.004737334791570902
step: 60, loss: 0.0006269204895943403
step: 70, loss: 0.005627214442938566
step: 80, loss: 0.04164453223347664
step: 90, loss: 0.0007645441801287234
step: 100, loss: 0.005962252151221037
step: 110, loss: 0.010965975001454353
step: 120, loss: 0.0025094039738178253
step: 130, loss: 0.027002187445759773
step: 140, loss: 0.003831072011962533
step: 150, loss: 0.007273488212376833
step: 160, loss: 0.005417186766862869
step: 170, loss: 0.00034177908673882484
step: 180, loss: 0.007122140843421221
step: 190, loss: 0.004748097155243158
/usr/local/lib/python3.7/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=0.7262872628726287, f1=0.7541899441340784,
best_f1=0.7679558011049724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026977861300110817
step: 10, loss: 0.0006210708525031805
step: 20, loss: 0.014987044036388397
step: 30, loss: 0.0011997671099379659
step: 40, loss: 0.005058846902102232
step: 50, loss: 0.007090857718139887
step: 60, loss: 0.0009789298055693507
step: 70, loss: 0.0025316462852060795
step: 80, loss: 0.02504110522568226
step: 90, loss: 0.0106001365929842
step: 100, loss: 0.01246155146509409
step: 110, loss: 0.007670190650969744
step: 120, loss: 0.047185178846120834
step: 130, loss: 0.002375558950006962
step: 140, loss: 0.0002320135827176273
step: 150, loss: 0.06653328984975815
step: 160, loss: 0.0005844978732056916
step: 170, loss: 0.07924990355968475
```

```
step: 180, loss: 0.1378643661737442
step: 190, loss: 0.0041544437408447266
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7525773195876289, f1=0.768421052631579,
best_f1=0.768421052631579
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005726313102059066
step: 10, loss: 0.0017825330141931772
step: 20, loss: 0.0004055197350680828
step: 30, loss: 0.0036108363419771194
step: 40, loss: 0.02950444445014
step: 50, loss: 0.0016593761974945664
step: 60, loss: 0.021228399127721786
step: 70, loss: 0.003382862778380513
step: 80, loss: 0.00018220565107185394
step: 90, loss: 0.008169317618012428
step: 100, loss: 0.000447084748884663
step: 110, loss: 0.0012727213324978948
step: 120, loss: 0.005618327762931585
step: 130, loss: 0.001957905013114214
step: 140, loss: 0.26124143600463867
step: 150, loss: 0.0005127902841195464
step: 160, loss: 0.0027207140810787678
step: 170, loss: 0.012830519117414951
step: 180, loss: 0.0009250693256035447
step: 190, loss: 0.0018783985869958997
/usr/local/lib/python3.7/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=0.7416267942583731, f1=0.7639902676399026,
best_f1=0.768421052631579
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001508016837760806
step: 10, loss: 0.001241647987626493
step: 20, loss: 0.05908886343240738
step: 30, loss: 0.0015811356715857983
step: 40, loss: 0.0003306951548438519
step: 50, loss: 0.0014879899099469185
```

```
step: 60, loss: 0.0008353898883797228
step: 70, loss: 0.00045620580203831196
step: 80, loss: 0.000609635841101408
step: 90, loss: 0.0038735661655664444
step: 100, loss: 0.00020268457592464983
step: 110, loss: 0.00016541904187761247
step: 120, loss: 0.016170088201761246
step: 130, loss: 0.0016593383625149727
step: 140, loss: 0.00034574023447930813
step: 150, loss: 0.001236405922099948
step: 160, loss: 0.00021283533715177327
step: 170, loss: 0.0007747574709355831
step: 180, loss: 0.007862097583711147
step: 190, loss: 0.00042764563113451004
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7543424317617866, f1=0.7645569620253165,
best_f1=0.7645569620253165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018836106755770743
step: 10, loss: 0.002358927857130766
step: 20, loss: 0.0008837394416332245
step: 30, loss: 0.0016413482371717691
step: 40, loss: 0.00042217239388264716
step: 50, loss: 0.002451512962579727
step: 60, loss: 0.0006581699126400054
step: 70, loss: 0.00025479099713265896
step: 80, loss: 0.0004004415823146701
step: 90, loss: 0.030180292204022408
step: 100, loss: 0.0002547868061810732
step: 110, loss: 0.0038285062182694674
step: 120, loss: 0.000441394658992067
step: 130, loss: 0.00014356253086589277
step: 140, loss: 0.00048763107042759657
step: 150, loss: 0.00022434306447394192
step: 160, loss: 0.002429886255413294
step: 170, loss: 0.004392955452203751
step: 180, loss: 0.0003998182946816087
step: 190, loss: 0.028722001239657402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7560975609756098, f1=0.7594936708860759,
best_f1=0.7594936708860759
/usr/local/lib/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 16384.0
step: 0, loss: 0.066155150532722247
step: 10, loss: 0.00021851836936548352
step: 20, loss: 0.0003879954165313393
step: 30, loss: 0.00028423263574950397
step: 40, loss: 0.00015730754239484668
step: 50, loss: 0.00030821614200249314
step: 60, loss: 0.00039929369813762605
step: 70, loss: 0.0038326415233314037
step: 80, loss: 0.00017179077258333564
step: 90, loss: 0.0021447085309773684
step: 100, loss: 0.0004091878072358668
step: 110, loss: 0.0016559686046093702
step: 120, loss: 0.00034654137562029064
step: 130, loss: 0.0006805927259847522
step: 140, loss: 0.00332739413715899
step: 150, loss: 0.0002218724985141307
step: 160, loss: 0.00020976191444788128
step: 170, loss: 0.0007541121449321508
step: 180, loss: 0.00021982108592055738
step: 190, loss: 0.0005391733138822019
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7582697201017812, f1=0.7624020887728459,
best_f1=0.7624020887728459
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003109200857579708
step: 10, loss: 0.0002194045518990606
step: 20, loss: 0.026390982791781425
step: 30, loss: 0.0006120107718743384
step: 40, loss: 0.00024801777908578515
```

```
step: 50, loss: 0.0002787260746117681
step: 60, loss: 0.00032397027825936675
step: 70, loss: 0.0004771998501382768
step: 80, loss: 9.219240746460855e-05
step: 90, loss: 0.00021795400243718177
step: 100, loss: 0.0020085787400603294
step: 110, loss: 0.00026348177925683356
step: 120, loss: 0.00016691276687197387
step: 130, loss: 0.00019880676700026794
step: 140, loss: 0.0008782631484791636
step: 150, loss: 0.00043249072041362524
step: 160, loss: 0.00031652781763114035
step: 170, loss: 0.00029073169571347535
step: 180, loss: 0.00022495018492918462
step: 190, loss: 0.00014984275912865996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7696335078534032, f1=0.7807486631016043,
best_f1=0.7807486631016043
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00032615402597002685
step: 10, loss: 0.00010375634883530438
step: 20, loss: 0.00034096461604349315
step: 30, loss: 0.00015557798906229436
step: 40, loss: 0.00044909046846441925
step: 50, loss: 0.019749440252780914
step: 60, loss: 0.00016051821876317263
step: 70, loss: 0.0005956100067123771
step: 80, loss: 0.016938868910074234
step: 90, loss: 0.0005387288401834667
step: 100, loss: 0.00037945719668641686
step: 110, loss: 0.00016025126387830824
step: 120, loss: 0.0001701127621345222
step: 130, loss: 0.0009887339547276497
step: 140, loss: 0.017609182745218277
step: 150, loss: 0.00032984971767291427
step: 160, loss: 0.0014054536586627364
step: 170, loss: 0.0013892815914005041
step: 180, loss: 0.00040513838757760823
step: 190, loss: 0.0024562524631619453
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.7708333333333334, f1=0.7733333333333333,
best_f1=0.7733333333333333
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 [00:05, 373.91it/s]
load_f1 = 0.7672634271099745
real_f1 = 0.7653061224489796
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 413.38it/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 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: 0.8470625877380371
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22721567749977112
step: 20, loss: 0.15505731105804443
step: 30, loss: 0.23151910305023193
step: 40, loss: 0.305692195892334
step: 50, loss: 0.3723002076148987
step: 60, loss: 0.4367283284664154
step: 70, loss: 0.30322086811065674
step: 80, loss: 0.26001492142677307
step: 90, loss: 0.38576453924179077
step: 100, loss: 0.2247529923915863
step: 110, loss: 0.179505854845047
step: 120, loss: 0.5554507374763489
step: 130, loss: 0.41611653566360474
step: 140, loss: 0.4665069580078125
step: 150, loss: 0.12031816691160202
step: 160, loss: 0.33832916617393494
step: 170, loss: 0.23096491396427155
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.365, f1=0.31876606683804626, best_f1=0.31876606683804626
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.370540052652359
step: 10, loss: 0.13761155307292938
step: 20, loss: 0.3248486816883087
step: 30, loss: 0.2415754795074463
step: 40, loss: 0.13457782566547394
step: 50, loss: 0.21202003955841064
step: 60, loss: 0.18610665202140808
step: 70, loss: 0.26406699419021606
step: 80, loss: 0.14671853184700012
step: 90, loss: 0.11131242662668228
step: 100, loss: 0.18887434899806976
step: 110, loss: 0.17070379853248596
step: 120, loss: 0.06987688690423965
step: 130, loss: 0.0343329980969429
step: 140, loss: 0.09375038743019104
step: 150, loss: 0.058107081800699234
step: 160, loss: 0.10900663584470749
step: 170, loss: 0.1810978204011917
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7945205479452053, f1=0.7602591792656588,
best_f1=0.7602591792656588
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04688238352537155
step: 10, loss: 0.033771440386772156
step: 20, loss: 0.046062976121902466
step: 30, loss: 0.17258356511592865
step: 40, loss: 0.027890149503946304
step: 50, loss: 0.11356745660305023
step: 60, loss: 0.1447754204273224
step: 70, loss: 0.04280328378081322
step: 80, loss: 0.13935257494449615
step: 90, loss: 0.09940747916698456
step: 100, loss: 0.03832694888114929
step: 110, loss: 0.03708716109395027
step: 120, loss: 0.002578380284830928
step: 130, loss: 0.1265052706003189
step: 140, loss: 0.007586750201880932
```

```
step: 150, loss: 0.03428972512483597
step: 160, loss: 0.04021435230970383
step: 170, loss: 0.20650453865528107
/usr/local/lib/python3.7/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=0.7695961995249406, f1=0.7852193995381063,
best_f1=0.7602591792656588
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03848964348435402
step: 10, loss: 0.1968885064125061
step: 20, loss: 0.023703977465629578
step: 30, loss: 0.04873046278953552
step: 40, loss: 0.0060637169517576694
step: 50, loss: 0.02900541014969349
step: 60, loss: 0.04097922518849373
step: 70, loss: 0.00717763788998127
step: 80, loss: 0.04555456340312958
step: 90, loss: 0.032790955156087875
step: 100, loss: 0.0076783361609578133
step: 110, loss: 0.0907670110464096
step: 120, loss: 0.11474540829658508
step: 130, loss: 0.04527677595615387
step: 140, loss: 0.011139264330267906
step: 150, loss: 0.0011329646222293377
step: 160, loss: 0.019233396276831627
step: 170, loss: 0.03961486369371414
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8049999999999999, f1=0.8201438848920863,
best_f1=0.8201438848920863
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020078308880329132
step: 10, loss: 0.009836978279054165
step: 20, loss: 0.007183381821960211
step: 30, loss: 0.02913912944495678
step: 40, loss: 0.013877426274120808
step: 50, loss: 0.031107131391763687
step: 60, loss: 0.005488743539899588
```

```
step: 70, loss: 0.01362366322427988
step: 80, loss: 0.0011360327480360866
step: 90, loss: 0.004258271306753159
step: 100, loss: 0.018301213160157204
step: 110, loss: 0.011238138191401958
step: 120, loss: 0.03824488818645477
step: 130, loss: 0.002749418141320348
step: 140, loss: 0.03748466074466705
step: 150, loss: 0.023587673902511597
step: 160, loss: 0.002466332633048296
step: 170, loss: 0.009615586139261723
/usr/local/lib/python3.7/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=0.7897196261682243, f1=0.7837837837837838,
best_f1=0.8201438848920863
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012691558338701725
step: 10, loss: 0.04988213628530502
step: 20, loss: 0.03683055564761162
step: 30, loss: 0.04654370993375778
step: 40, loss: 0.04978625476360321
step: 50, loss: 0.018694689497351646
step: 60, loss: 0.00786988902837038
step: 70, loss: 0.008830861188471317
step: 80, loss: 0.02550445683300495
step: 90, loss: 0.00352028151974082
step: 100, loss: 0.04724189639091492
step: 110, loss: 0.004654783755540848
step: 120, loss: 0.017287205904722214
step: 130, loss: 0.18286024034023285
step: 140, loss: 0.06794444471597672
step: 150, loss: 0.10527047514915466
step: 160, loss: 0.02272619865834713
step: 170, loss: 0.0010857598390430212
/usr/local/lib/python3.7/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=0.7958656330749355, f1=0.801980198019802,
best_f1=0.8201438848920863
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0034454679116606712
step: 10, loss: 0.000656493182759732
step: 20, loss: 0.0015401310520246625
```

```
step: 30, loss: 0.0009316214127466083
step: 40, loss: 0.012784653343260288
step: 50, loss: 0.0024708949495106936
step: 60, loss: 0.08124513924121857
step: 70, loss: 0.0011943416902795434
step: 80, loss: 0.0011907980078831315
step: 90, loss: 0.011195178143680096
step: 100, loss: 0.02757023461163044
step: 110, loss: 0.02198505587875843
step: 120, loss: 0.1120796725153923
step: 130, loss: 0.03656192496418953
step: 140, loss: 0.0010908080730587244
step: 150, loss: 0.021017147228121758
step: 160, loss: 0.0035725929774343967
step: 170, loss: 0.0020131669007241726
/usr/local/lib/python3.7/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=0.7720207253886009, f1=0.792079207920792,
best_f1=0.8201438848920863
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01296076737344265
step: 10, loss: 0.0009765063878148794
step: 20, loss: 0.006708640605211258
step: 30, loss: 0.014292960986495018
step: 40, loss: 0.0005608224892057478
step: 50, loss: 0.000606756133493036
step: 60, loss: 0.06855863332748413
step: 70, loss: 0.020229890942573547
step: 80, loss: 0.0019367834320291877
step: 90, loss: 0.009940685704350471
step: 100, loss: 0.00493084779009223
step: 110, loss: 0.0035816156305372715
step: 120, loss: 0.0016069673001766205
step: 130, loss: 0.007038759998977184
step: 140, loss: 0.00045737193431705236
step: 150, loss: 0.04110942780971527
step: 160, loss: 0.025259658694267273
step: 170, loss: 0.0022810243535786867
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8121827411167513, f1=0.8289156626506025,
best_f1=0.8289156626506025
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004278920125216246
step: 10, loss: 0.0042294398881495
step: 20, loss: 0.0005174722173251212
step: 30, loss: 0.00974796712398529
step: 40, loss: 0.002087772823870182
step: 50, loss: 0.016056397929787636
step: 60, loss: 0.00041276286356151104
step: 70, loss: 0.006375557277351618
step: 80, loss: 0.02431604452431202
step: 90, loss: 0.0030912442598491907
step: 100, loss: 0.01699884608387947
step: 110, loss: 9.151938138529658e-05
step: 120, loss: 0.0006542187184095383
step: 130, loss: 0.0007356959977187216
step: 140, loss: 0.0036473635118454695
step: 150, loss: 0.014179293066263199
step: 160, loss: 0.0029852967709302902
step: 170, loss: 0.008034991100430489
/usr/local/lib/python3.7/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=0.8042328042328043, f1=0.8140703517587939,
best_f1=0.8289156626506025
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006180680356919765
step: 10, loss: 0.006264344323426485
step: 20, loss: 0.0005957381799817085
step: 30, loss: 0.0005090293707326055
step: 40, loss: 0.0037561177741736174
step: 50, loss: 0.0032224967144429684
step: 60, loss: 0.00033526099286973476
step: 70, loss: 0.00024040341668296605
step: 80, loss: 0.10056290775537491
step: 90, loss: 0.028568783774971962
step: 100, loss: 0.006007405463606119
step: 110, loss: 0.0005635148845613003
step: 120, loss: 0.013656094670295715
step: 130, loss: 0.0029912276659160852
step: 140, loss: 0.0037180865183472633
step: 150, loss: 0.000347523222444579
step: 160, loss: 7.132127939257771e-05
```

```
step: 170, loss: 0.0003188755363225937
/usr/local/lib/python3.7/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=0.8121827411167513, f1=0.8280871670702179,
best_f1=0.8289156626506025
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00035776346339844167
step: 10, loss: 0.007940331473946571
step: 20, loss: 0.006961758714169264
step: 30, loss: 0.0007858941680751741
step: 40, loss: 0.013402191922068596
step: 50, loss: 0.0006540751783177257
step: 60, loss: 0.004554762504994869
step: 70, loss: 0.00023064135166350752
step: 80, loss: 0.003884960198774934
step: 90, loss: 0.009058096446096897
step: 100, loss: 0.03886597976088524
step: 110, loss: 0.0010179259115830064
step: 120, loss: 0.0007396408473141491
step: 130, loss: 0.00025760300923138857
step: 140, loss: 0.006495581939816475
step: 150, loss: 0.0008323374786414206
step: 160, loss: 0.0008187966886907816
step: 170, loss: 0.0445445217192173
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8214285714285714, f1=0.8235294117647058,
best_f1=0.8235294117647058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010299600660800934
step: 10, loss: 0.000491796643473208
step: 20, loss: 0.049661155790009056
step: 30, loss: 0.0022165016271173954
step: 40, loss: 0.0002970848581753671
step: 50, loss: 0.0002523388247936964
step: 60, loss: 0.0035021291114389896
step: 70, loss: 0.008478687144815922
step: 80, loss: 0.109534353017807
```

```
step: 90, loss: 0.00037693014019168913
step: 100, loss: 0.00920170359313488
step: 110, loss: 0.00013888490502722561
step: 120, loss: 0.00031627333373762667
step: 130, loss: 0.0010859703179448843
step: 140, loss: 0.000464361859485507
step: 150, loss: 0.0002722966019064188
step: 160, loss: 0.006582265719771385
step: 170, loss: 0.001957378350198269
/usr/local/lib/python3.7/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=0.8167539267015707, f1=0.8232323232323232,
best_f1=0.8235294117647058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00045025750296190038
step: 10, loss: 0.035007115453481674
step: 20, loss: 0.0007924222154542804
step: 30, loss: 0.0003661189693957567
step: 40, loss: 0.0010236361995339394
step: 50, loss: 0.0005040069227106869
step: 60, loss: 0.0027142027392983437
step: 70, loss: 0.0025514804292470217
step: 80, loss: 0.004547876305878162
step: 90, loss: 0.0001823514758143574
step: 100, loss: 0.0065691485069692135
step: 110, loss: 0.00026169620105065405
step: 120, loss: 0.0002046803419943899
step: 130, loss: 0.005242407321929932
step: 140, loss: 9.708775905892253e-05
step: 150, loss: 0.00028799567371606827
step: 160, loss: 0.00030201749177649617
step: 170, loss: 0.0038058722857385874
/usr/local/lib/python3.7/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=0.8165374677002585, f1=0.8229426433915212,
best_f1=0.8235294117647058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00019005774811375886
step: 10, loss: 0.0004751983215101063
step: 20, loss: 0.0003811008937191218
step: 30, loss: 0.012107886373996735
step: 40, loss: 0.00026483822148293257
```

```
step: 50, loss: 0.0022272276692092242
step: 60, loss: 0.00024441556888632476
step: 70, loss: 0.0035108504816889763
step: 80, loss: 0.0010061623761430383
step: 90, loss: 0.00011237381841056049
step: 100, loss: 0.00017469703743699938
step: 110, loss: 0.00091636145953089
step: 120, loss: 0.0092853352192819118
step: 130, loss: 0.00016825842612888664
step: 140, loss: 0.0026077425573021173
step: 150, loss: 0.0002024790592258796
step: 160, loss: 0.04182325303554535
step: 170, loss: 0.0007185464492067695
/usr/local/lib/python3.7/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=0.8207792207792208, f1=0.8270676691729323,
best_f1=0.8235294117647058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016970142314676195
step: 10, loss: 0.008857529610395432
step: 20, loss: 0.0013839594321325421
step: 30, loss: 0.08533627539873123
step: 40, loss: 9.385775774717331e-05
step: 50, loss: 0.0014466335996985435
step: 60, loss: 7.23703415133059e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.10137603431940079
step: 80, loss: 0.0628967434167862
step: 90, loss: 0.0003455407568253577
step: 100, loss: 0.0014596605906262994
step: 110, loss: 0.00026046574930660427
step: 120, loss: 0.003804263658821583
step: 130, loss: 0.0006529103848151863
step: 140, loss: 0.00018858889234252274
step: 150, loss: 0.03813133016228676
step: 160, loss: 0.00010482077777851373
step: 170, loss: 0.0008954308577813208
/usr/local/lib/python3.7/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=0.8184143222506394, f1=0.819753086419753,
best_f1=0.8235294117647058
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 [00:04, 427.69it/s]
load_f1 = 0.8181818181818181
real_f1 = 0.8190954773869347
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, 364.56it/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: 0.8165208697319031
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45675116777420044
step: 20, loss: 0.5832545161247253
step: 30, loss: 0.41203591227531433
step: 40, loss: 0.15800438821315765
step: 50, loss: 0.054438356310129166
step: 60, loss: 0.07726985216140747
step: 70, loss: 0.11674748361110687
step: 80, loss: 0.20727238059043884
step: 90, loss: 0.02555697225034237
step: 100, loss: 0.050054263323545456
step: 110, loss: 0.13216844201087952
step: 120, loss: 0.02455969899892807
step: 130, loss: 0.012318785302340984
step: 140, loss: 0.007837090641260147
step: 150, loss: 0.17168670892715454
step: 160, loss: 0.19115447998046875
step: 170, loss: 0.05812555551528931
step: 180, loss: 0.010842902585864067
step: 190, loss: 0.012973848730325699
step: 200, loss: 0.00741164293140173
step: 210, loss: 0.015151793137192726
step: 220, loss: 0.008250265382230282
step: 230, loss: 0.025977272540330887
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9664429530201343, f1=0.9605411499436302,
best_f1=0.9605411499436302
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02063043788075447

```
step: 10, loss: 0.02648509480059147
step: 20, loss: 0.0024600857868790627
step: 30, loss: 0.0030174902640283108
step: 40, loss: 0.015270697884261608
step: 50, loss: 0.016159096732735634
step: 60, loss: 0.005096482113003731
step: 70, loss: 0.0164360161870718
step: 80, loss: 0.0015394918154925108
step: 90, loss: 0.012264047749340534
step: 100, loss: 0.1324017345905304
step: 110, loss: 0.1293756067752838
step: 120, loss: 0.013397152535617352
step: 130, loss: 0.0032308497466146946
step: 140, loss: 0.2751934826374054
step: 150, loss: 0.013038038276135921
step: 160, loss: 0.008861664682626724
step: 170, loss: 0.020564809441566467
step: 180, loss: 0.023837460204958916
step: 190, loss: 0.21690483391284943
step: 200, loss: 0.030969170853495598
step: 210, loss: 0.028512578457593918
step: 220, loss: 0.008120189420878887
step: 230, loss: 0.010728619061410427
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9776286353467561, f1=0.967741935483871,
best_f1=0.967741935483871
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01918683759868145
step: 10, loss: 0.0130311269313097
step: 20, loss: 0.003463038709014654
step: 30, loss: 0.008238544687628746
step: 40, loss: 0.014029797166585922
step: 50, loss: 0.0037013643886893988
step: 60, loss: 0.014480960555374622
step: 70, loss: 0.006955751217901707
step: 80, loss: 0.0035526822321116924
step: 90, loss: 0.09508628398180008
step: 100, loss: 0.006394860800355673
step: 110, loss: 0.01170277874916792
step: 120, loss: 0.002899925922974944
```

```
step: 130, loss: 0.012575767934322357
step: 140, loss: 0.0008370562572963536
step: 150, loss: 0.0009645713726058602
step: 160, loss: 0.0027696480974555016
step: 170, loss: 0.011791860684752464
step: 180, loss: 0.01185125857591629
step: 190, loss: 0.006450612563639879
step: 200, loss: 0.015097920782864094
step: 210, loss: 0.018304964527487755
step: 220, loss: 0.006074569653719664
step: 230, loss: 0.012120725587010384
/usr/local/lib/python3.7/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=0.9731543624161074, f1=0.9685393258426966,
best_f1=0.967741935483871
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0075265197083354
step: 10, loss: 0.001932397368364036
step: 20, loss: 0.001212284667417407
step: 30, loss: 0.019282912835478783
step: 40, loss: 0.0031675007194280624
step: 50, loss: 0.002205935539677739
step: 60, loss: 0.0029753223061561584
step: 70, loss: 0.0071296109817922115
step: 80, loss: 0.10098199546337128
step: 90, loss: 0.07988007366657257
step: 100, loss: 0.002096950775012374
step: 110, loss: 0.013048414140939713
step: 120, loss: 0.0033911701757460833
step: 130, loss: 0.001589482999406755
step: 140, loss: 0.0004126516869291663
step: 150, loss: 0.02738834358751774
step: 160, loss: 0.0006122879567556083
step: 170, loss: 0.009798107668757439
step: 180, loss: 0.14350757002830505
step: 190, loss: 0.002513349987566471
step: 200, loss: 0.0018683611415326595
step: 210, loss: 0.0016154705081135035
step: 220, loss: 0.0011741627240553498
step: 230, loss: 0.0017205196199938655
/usr/local/lib/python3.7/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=0.9762174405436014, f1=0.9647326507394768,
best_f1=0.967741935483871
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011283979984000325
step: 10, loss: 0.0015523293986916542
step: 20, loss: 0.0032654856331646442
step: 30, loss: 0.0004890933050774038
step: 40, loss: 0.0014551331987604499
step: 50, loss: 0.0007473933510482311
step: 60, loss: 0.0009815379744395614
step: 70, loss: 0.00028863383340649307
step: 80, loss: 0.0005801749066449702
step: 90, loss: 0.0006842619623057544
step: 100, loss: 0.0022061604540795088
step: 110, loss: 0.008069208823144436
step: 120, loss: 0.026156041771173477
step: 130, loss: 0.038337342441082
step: 140, loss: 0.0011866516433656216
step: 150, loss: 0.0004886511596851051
step: 160, loss: 0.12401992082595825
step: 170, loss: 0.06832213699817657
step: 180, loss: 0.0020452619064599276
step: 190, loss: 0.001101854257285595
step: 200, loss: 0.008062650449573994
step: 210, loss: 0.0014797524781897664
step: 220, loss: 0.00036418679519556463
step: 230, loss: 0.001753546530380845
/usr/local/lib/python3.7/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=0.9741282339707535, f1=0.9707207207207207,
best_f1=0.967741935483871
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007624171674251556
step: 10, loss: 0.00029112875927239656
step: 20, loss: 0.0009064107434824109
step: 30, loss: 0.005064213182777166
step: 40, loss: 0.0028499497566372156
step: 50, loss: 0.07600708305835724
step: 60, loss: 0.02400710992515087
step: 70, loss: 0.0031732236966490746
step: 80, loss: 0.0008664206834509969
step: 90, loss: 0.0006942130858078599
step: 100, loss: 0.0010852162959054112
step: 110, loss: 0.017734695225954056
step: 120, loss: 0.0006885526818223298
```

```
step: 130, loss: 0.000488328339997679
step: 140, loss: 0.001759809092618525
step: 150, loss: 0.03263305127620697
step: 160, loss: 0.06818149238824844
step: 170, loss: 0.0006133491988293827
step: 180, loss: 0.003991981036961079
step: 190, loss: 0.001638853456825018
step: 200, loss: 0.001599297160282731
step: 210, loss: 0.0016972713638097048
step: 220, loss: 0.0008467993466183543
step: 230, loss: 0.0005325753008946776
```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.972129319955407, f1=0.9721913236929923,
best_f1=0.967741935483871
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```
step: 0, loss: 0.10367432981729507
step: 10, loss: 0.0016890509286895394
step: 20, loss: 0.00038685143226757646
step: 30, loss: 0.0008057865779846907
step: 40, loss: 0.00034161205985583365
step: 50, loss: 0.000468511541839689
step: 60, loss: 0.01607188954949379
step: 70, loss: 0.001014566165395081
step: 80, loss: 0.00016766611952334642
step: 90, loss: 0.0005319127812981606
step: 100, loss: 0.0002942102146334946
step: 110, loss: 0.003949536941945553
step: 120, loss: 0.002679545432329178
step: 130, loss: 0.007355187553912401
step: 140, loss: 0.0003381569404155016
step: 150, loss: 0.001080034300684929
step: 160, loss: 0.00021654345619026572
step: 170, loss: 0.0001600707764737308
step: 180, loss: 0.0004868711985182017
step: 190, loss: 0.0009225660469383001
step: 200, loss: 0.004063721280544996
step: 210, loss: 0.0010455051669850945
step: 220, loss: 0.001274276408366859
step: 230, loss: 0.02130158804357052
```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9789590254706534, f1=0.9713656387665198,
best_f1=0.9713656387665198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0094261157921552658
step: 10, loss: 0.01617520861327648
step: 20, loss: 0.0003943739284295589
step: 30, loss: 0.00026812113355845213
step: 40, loss: 0.0006522256880998611
step: 50, loss: 0.000540341658052057
step: 60, loss: 0.0003138415631838143
step: 70, loss: 0.003561581950634718
step: 80, loss: 0.001525728264823556
step: 90, loss: 0.00022976353648118675
step: 100, loss: 0.00255592935718596
step: 110, loss: 0.00024950108490884304
step: 120, loss: 0.00020067549485247582
step: 130, loss: 0.0020158844999969006
step: 140, loss: 0.0011867438443005085
step: 150, loss: 0.00025305384770035744
step: 160, loss: 0.0007330991793423891
step: 170, loss: 0.0001222570426762104
step: 180, loss: 0.0013287608744576573
step: 190, loss: 0.00011900547542609274
step: 200, loss: 0.0009639210184104741
step: 210, loss: 0.0002706532832235098
step: 220, loss: 0.00019627399160526693
step: 230, loss: 0.09952207654714584
/usr/local/lib/python3.7/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=0.9766925638179801, f1=0.9702970297029702,
best_f1=0.9713656387665198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.656233305577189e-05
step: 10, loss: 0.00030142502509988844
step: 20, loss: 0.0011520225089043379
step: 30, loss: 0.0015635554445907474
step: 40, loss: 0.000148427949170582
step: 50, loss: 0.0014683292247354984
step: 60, loss: 0.00048659404274076223
step: 70, loss: 0.0004184653516858816
step: 80, loss: 0.029236508533358574

```
step: 90, loss: 0.01295443158596754
step: 100, loss: 0.004400898702442646
step: 110, loss: 0.000372438138583675
step: 120, loss: 0.029043953865766525
step: 130, loss: 0.000110982233309187
step: 140, loss: 0.004289079923182726
step: 150, loss: 5.8893168898066506e-05
step: 160, loss: 0.00020117312669754028
step: 170, loss: 6.798382673878223e-05
step: 180, loss: 0.00021163855853956193
step: 190, loss: 0.002243733499199152
step: 200, loss: 0.0006341270636767149
step: 210, loss: 0.00023882862296886742
step: 220, loss: 0.17447420954704285
step: 230, loss: 0.0008407820132561028
/usr/local/lib/python3.7/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=0.9765886287625419, f1=0.9755555555555556,
best_f1=0.9713656387665198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012677829363383353
step: 10, loss: 8.022107795113698e-05
step: 20, loss: 0.0008074697107076645
step: 30, loss: 0.000172135783941485
step: 40, loss: 0.00014327729877550155
step: 50, loss: 0.0004883428919129074
step: 60, loss: 0.05382388085126877
step: 70, loss: 0.00016763580788392574
step: 80, loss: 0.00016251821944024414
step: 90, loss: 0.0005457791849039495
step: 100, loss: 0.000720027950592339
step: 110, loss: 0.0014137333491817117
step: 120, loss: 0.009150351397693157
step: 130, loss: 0.0003422195972688496
step: 140, loss: 0.017776969820261
step: 150, loss: 8.60266518429853e-05
step: 160, loss: 7.640212425030768e-05
step: 170, loss: 0.0003297080402262509
step: 180, loss: 0.00038727145874872804
step: 190, loss: 0.00124509003944695
step: 200, loss: 6.985642539802939e-05
step: 210, loss: 0.00011090299813076854
step: 220, loss: 0.00017115619266405702
step: 230, loss: 0.0009111062390729785
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9797752808988766, f1=0.9752252252252253,
best_f1=0.9752252252252253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.699285146780312e-05
step: 10, loss: 0.00011621154408203438
step: 20, loss: 3.76871794287581e-05
step: 30, loss: 9.21053797355853e-05
step: 40, loss: 0.0002474718203302473
step: 50, loss: 0.0002703357022255659
step: 60, loss: 7.924811507109553e-05
step: 70, loss: 0.001130619435571134
step: 80, loss: 0.0006365043227560818
step: 90, loss: 0.0015647037653252482
step: 100, loss: 0.00019013947166968137
step: 110, loss: 8.399791113333777e-05
step: 120, loss: 0.0001210862465086393
step: 130, loss: 4.3967258534394205e-05
step: 140, loss: 0.0002368828427279368
step: 150, loss: 6.620914064114913e-05
step: 160, loss: 0.0003707975265569985
step: 170, loss: 0.0038610941264778376
step: 180, loss: 0.00021092875977046788
step: 190, loss: 0.00018392295169178396
step: 200, loss: 0.00020898380898870528
step: 210, loss: 6.325372669380158e-05
step: 220, loss: 7.367619400611147e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 230, loss: 0.023929785937070847
/usr/local/lib/python3.7/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=0.9776785714285714, f1=0.9764837625979844,
best_f1=0.9752252252252253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012187986430944875
step: 10, loss: 8.625539339846e-05
step: 20, loss: 5.470210817293264e-05
step: 30, loss: 0.0031556340400129557
```

```
step: 40, loss: 6.444675091188401e-05
step: 50, loss: 9.421911818208173e-05
step: 60, loss: 0.0003566366503946483
step: 70, loss: 0.00012816568778362125
step: 80, loss: 3.92553374695126e-05
step: 90, loss: 0.016212420538067818
step: 100, loss: 8.898235682863742e-05
step: 110, loss: 0.00033292235457338393
step: 120, loss: 8.3871120295953e-05
step: 130, loss: 0.0030954605899751186
step: 140, loss: 4.721934601548128e-05
step: 150, loss: 5.269028406473808e-05
step: 160, loss: 0.0294507946819067
step: 170, loss: 7.341773743974045e-05
step: 180, loss: 7.268170884344727e-05
step: 190, loss: 0.00010550102160777897
step: 200, loss: 0.003492089884233475
step: 210, loss: 9.808203321881592e-05
step: 220, loss: 0.0005218064761720598
step: 230, loss: 6.573497375939041e-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=0.978675645342312, f1=0.9742441209406495,
best_f1=0.9752252252252253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.274662104668096e-05
step: 10, loss: 0.00012972358672413975
step: 20, loss: 8.69914292707108e-05
step: 30, loss: 0.014820271171629429
step: 40, loss: 9.985898941522464e-05
step: 50, loss: 8.340565545950085e-05
step: 60, loss: 4.0488375816494226e-05
step: 70, loss: 5.5485015764134005e-05
step: 80, loss: 4.560450179269537e-05
step: 90, loss: 3.2442694646306336e-05
step: 100, loss: 8.842381794238463e-05
step: 110, loss: 5.2587027312256396e-05
step: 120, loss: 6.861707515781745e-05
step: 130, loss: 7.363082113442943e-05
step: 140, loss: 0.0002579836000222713
step: 150, loss: 0.003501863218843937
step: 160, loss: 9.002235310617834e-05
step: 170, loss: 7.434000872308388e-05
step: 180, loss: 0.00017434079200029373
step: 190, loss: 6.487949576694518e-05
```

```
step: 200, loss: 4.610152245732024e-05
step: 210, loss: 6.62962265778333e-05
step: 220, loss: 4.856274608755484e-05
step: 230, loss: 4.0066937799565494e-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=0.9765363128491621, f1=0.9766407119021134,
best_f1=0.9752252252252253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.42422164673917e-05
step: 10, loss: 3.0032486392883584e-05
step: 20, loss: 9.165995288640261e-05
step: 30, loss: 7.027172978268936e-05
step: 40, loss: 3.9996037230594084e-05
step: 50, loss: 0.00031356720137409866
step: 60, loss: 3.55221563950181e-05
step: 70, loss: 7.118221401469782e-05
step: 80, loss: 5.884770507691428e-05
step: 90, loss: 0.0001655530941206962
step: 100, loss: 0.0071042305789887905
step: 110, loss: 0.00010325407492928207
step: 120, loss: 4.3199925130466e-05
step: 130, loss: 7.26111902622506e-05
step: 140, loss: 3.714310878422111e-05
step: 150, loss: 6.16888573858887e-05
step: 160, loss: 3.6882080166833475e-05
step: 170, loss: 6.534343265229836e-05
step: 180, loss: 6.845722964499146e-05
step: 190, loss: 0.0024216698948293924
step: 200, loss: 3.90428576793056e-05
step: 210, loss: 7.928556442493573e-05
step: 220, loss: 0.00011891961185028777
step: 230, loss: 2.5223394914064556e-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=0.9787234042553192, f1=0.9776785714285714,
best_f1=0.9752252252252253
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.982669608783908e-05
step: 10, loss: 4.835108484257944e-05
step: 20, loss: 8.69917421368882e-05
step: 30, loss: 6.407633918570355e-05
```

```
step: 40, loss: 6.831808423157781e-05
step: 50, loss: 7.106423436198384e-05
step: 60, loss: 3.203653614036739e-05
step: 70, loss: 6.460196163970977e-05
step: 80, loss: 0.003874561982229352
step: 90, loss: 2.973442133225035e-05
step: 100, loss: 5.779792263638228e-05
step: 110, loss: 6.576922169188038e-05
step: 120, loss: 5.9636779042193666e-05
step: 130, loss: 7.718660344835371e-05
step: 140, loss: 0.023431723937392235
step: 150, loss: 7.497149636037648e-05
step: 160, loss: 8.299265755340457e-05
step: 170, loss: 3.969469617004506e-05
step: 180, loss: 9.644885722082108e-05
step: 190, loss: 0.00012271708692424
step: 200, loss: 4.363167317933403e-05
step: 210, loss: 0.0010905630188062787
step: 220, loss: 0.0004887668183073401
step: 230, loss: 6.026733171893284e-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=0.9775280898876404, f1=0.9730941704035874,
best_f1=0.9752252252252253
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
```

```
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 313.35it/s]
load_f1 = 0.9797752808988766
real_f1 = 0.9775280898876404
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 [00:13, 332.72it/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 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: 0.7939895391464233
step: 10, loss: 0.4022519886493683
step: 20, loss: 0.4936389625072479
step: 30, loss: 0.42556285858154297
step: 40, loss: 0.3439832627773285
step: 50, loss: 0.17551937699317932
step: 60, loss: 0.18642094731330872
step: 70, loss: 0.07195276021957397
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.11133790761232376
step: 90, loss: 0.13337743282318115
step: 100, loss: 0.30682605504989624
step: 110, loss: 0.05359937995672226

```
step: 120, loss: 0.032724276185035706
step: 130, loss: 0.06101558357477188
step: 140, loss: 0.1480000764131546
step: 150, loss: 0.020417345687747
step: 160, loss: 0.1821741759777069
step: 170, loss: 0.1579587310552597
step: 180, loss: 0.06844894587993622
step: 190, loss: 0.03393226116895676
step: 200, loss: 0.12711969017982483
step: 210, loss: 0.09642930328845978
step: 220, loss: 0.108220674097538
step: 230, loss: 0.11280949413776398
step: 240, loss: 0.035183317959308624
step: 250, loss: 0.018749352544546127
step: 260, loss: 0.04095572233200073
step: 270, loss: 0.014056386426091194
step: 280, loss: 0.10940532386302948
step: 290, loss: 0.07714460790157318
step: 300, loss: 0.15709327161312103
step: 310, loss: 0.08426765352487564
step: 320, loss: 0.024812472984194756
step: 330, loss: 0.05712650343775749
step: 340, loss: 0.1349666863679886
step: 350, loss: 0.05834762379527092
step: 360, loss: 0.04113991931080818
step: 370, loss: 0.08532138168811798
step: 380, loss: 0.10327178239822388
step: 390, loss: 0.013878308236598969
step: 400, loss: 0.005012174602597952
step: 410, loss: 0.02946876920759678
step: 420, loss: 0.040416419506073
step: 430, loss: 0.06134423986077309
step: 440, loss: 0.06267019361257553
step: 450, loss: 0.030795743688941002
step: 460, loss: 0.113261878490448
step: 470, loss: 0.1591891646385193
step: 480, loss: 0.35884442925453186
step: 490, loss: 0.029835164546966553
step: 500, loss: 0.017758270725607872
step: 510, loss: 0.06338122487068176
step: 520, loss: 0.11831702291965485
step: 530, loss: 0.11271245777606964
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9332723948811701, f1=0.9307201458523247,
best_f1=0.9307201458523247
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09642824530601501
step: 10, loss: 0.19795817136764526
step: 20, loss: 0.04312964901328087
step: 30, loss: 0.024682026356458664
step: 40, loss: 0.004834250081330538
step: 50, loss: 0.19069434702396393
step: 60, loss: 0.1218179389834404
step: 70, loss: 0.3069348931312561
step: 80, loss: 0.03424901142716408
step: 90, loss: 0.004748151637613773
step: 100, loss: 0.22517821192741394
step: 110, loss: 0.05480828881263733
step: 120, loss: 0.06518533825874329
step: 130, loss: 0.012076477520167828
step: 140, loss: 0.057304248213768005
step: 150, loss: 0.0707269236445427
step: 160, loss: 0.024689987301826477
step: 170, loss: 0.15854519605636597
step: 180, loss: 0.020057862624526024
step: 190, loss: 0.029936086386442184
step: 200, loss: 0.09753276407718658
step: 210, loss: 0.0017998580588027835
step: 220, loss: 0.16047318279743195
step: 230, loss: 0.026651497930288315
step: 240, loss: 0.13054576516151428
step: 250, loss: 0.060169704258441925
step: 260, loss: 0.034445326775312424
step: 270, loss: 0.09798972308635712
step: 280, loss: 0.08311167359352112
step: 290, loss: 0.06347797065973282
step: 300, loss: 0.010180709883570671
step: 310, loss: 0.04802617430686951
step: 320, loss: 0.09531494230031967
step: 330, loss: 0.015667445957660675
step: 340, loss: 0.02663002721965313
step: 350, loss: 0.02767682820558548
step: 360, loss: 0.010678923688828945
step: 370, loss: 0.0040759616531431675
step: 380, loss: 0.12586839497089386
step: 390, loss: 0.04016720503568649
step: 400, loss: 0.04511650279164314
step: 410, loss: 0.0004226212331559509
```

```
step: 420, loss: 0.04133787006139755
step: 430, loss: 0.020695380866527557
step: 440, loss: 0.012288020923733711
step: 450, loss: 0.08368989080190659
step: 460, loss: 0.260057657957077
step: 470, loss: 0.009433791041374207
step: 480, loss: 0.15743322670459747
step: 490, loss: 0.02360720746219158
step: 500, loss: 0.008004916831851006
step: 510, loss: 0.06114041060209274
step: 520, loss: 0.18559575080871582
step: 530, loss: 0.10836860537528992
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9464285714285713, f1=0.9343955014058105,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008935763500630856
step: 10, loss: 0.03418535366654396
step: 20, loss: 0.15835288166999817
step: 30, loss: 0.26249203085899353
step: 40, loss: 0.0037429966032505035
step: 50, loss: 0.0077859084121882915
step: 60, loss: 0.03716785088181496
step: 70, loss: 0.024253668263554573
step: 80, loss: 0.0037006770726293325
step: 90, loss: 0.012393748387694359
step: 100, loss: 0.036579158157110214
step: 110, loss: 0.00648775789886713
step: 120, loss: 0.005372190847992897
step: 130, loss: 0.02160806953907013
step: 140, loss: 0.017271939665079117
step: 150, loss: 0.0021603500936180353
step: 160, loss: 0.0016065459931269288
step: 170, loss: 0.009129959158599377
step: 180, loss: 0.003360806265845895
step: 190, loss: 0.0654086172580719
step: 200, loss: 0.0028612581081688404
step: 210, loss: 0.04936279356479645
step: 220, loss: 0.010927945375442505
step: 230, loss: 0.047266557812690735
```

```
step: 240, loss: 0.023565277457237244
step: 250, loss: 0.029462134465575218
step: 260, loss: 0.0006437803385779262
step: 270, loss: 0.0010957878548651934
step: 280, loss: 0.013882185332477093
step: 290, loss: 0.0572359561920166
step: 300, loss: 0.2145012766122818
step: 310, loss: 0.14440426230430603
step: 320, loss: 0.12625689804553986
step: 330, loss: 0.0018400625558570027
step: 340, loss: 0.0013125749537721276
step: 350, loss: 0.0034063835628330708
step: 360, loss: 0.022959725931286812
step: 370, loss: 0.001023122575134039
step: 380, loss: 0.014435011893510818
step: 390, loss: 0.013949431478977203
step: 400, loss: 0.018237173557281494
step: 410, loss: 0.00657917745411396
step: 420, loss: 0.011273822747170925
step: 430, loss: 0.15988831222057343
step: 440, loss: 0.025900883600115776
step: 450, loss: 0.05247780308127403
step: 460, loss: 0.0955352857708931
step: 470, loss: 0.004922181833535433
step: 480, loss: 0.006684770341962576
step: 490, loss: 0.13160844147205353
step: 500, loss: 0.036029152572155
step: 510, loss: 0.005714080296456814
step: 520, loss: 0.0023780204355716705
step: 530, loss: 0.05208062380552292
/usr/local/lib/python3.7/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=0.9410138248847926, f1=0.9413936317489617,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006420332007110119
step: 10, loss: 0.0034061018377542496
step: 20, loss: 0.04102249816060066
step: 30, loss: 0.016392052173614502
step: 40, loss: 0.006168219726532698
step: 50, loss: 0.06127721071243286
step: 60, loss: 0.0002373357128817588
step: 70, loss: 0.0027077484410256147
step: 80, loss: 0.014542704448103905
step: 90, loss: 0.0004136128700338304
```

```
step: 100, loss: 0.0004910845891572535
step: 110, loss: 0.00027322384994477034
step: 120, loss: 0.0020471138413995504
step: 130, loss: 0.07940971106290817
step: 140, loss: 0.23332099616527557
step: 150, loss: 0.012618470937013626
step: 160, loss: 0.011777913197875023
step: 170, loss: 0.0026325921062380075
step: 180, loss: 0.007500193547457457
step: 190, loss: 0.004653808660805225
step: 200, loss: 0.00039444261346943676
step: 210, loss: 0.10453664511442184
step: 220, loss: 0.0007937199552543461
step: 230, loss: 0.25446537137031555
step: 240, loss: 0.00913454033434391
step: 250, loss: 0.04443003982305527
step: 260, loss: 0.12705379724502563
step: 270, loss: 0.002449135296046734
step: 280, loss: 0.006409328430891037
step: 290, loss: 0.05478186160326004
step: 300, loss: 0.0013060718774795532
step: 310, loss: 0.002203794429078698
step: 320, loss: 0.00758361117914319
step: 330, loss: 0.14458899199962616
step: 340, loss: 0.006009347271174192
step: 350, loss: 0.0007699390989728272
step: 360, loss: 0.021478045731782913
step: 370, loss: 0.03588765859603882
step: 380, loss: 0.0006347776507027447
step: 390, loss: 0.04516041278839111
step: 400, loss: 0.018171876668930054
step: 410, loss: 0.011888829991221428
step: 420, loss: 0.00893343985080719
step: 430, loss: 0.06016033515334129
step: 440, loss: 0.09207545965909958
step: 450, loss: 0.03774171322584152
step: 460, loss: 0.0030475661624222994
step: 470, loss: 0.015736212953925133
step: 480, loss: 0.006609591189771891
step: 490, loss: 0.011537023819983006
step: 500, loss: 0.0057196421548724174
step: 510, loss: 0.014376213774085045
step: 520, loss: 0.026837369427084923
step: 530, loss: 0.003308321116492152
/usr/local/lib/python3.7/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=0.9382600561272217, f1=0.9387947269303202,
```

```
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0030644163489341736
step: 10, loss: 0.062875382602221481
step: 20, loss: 0.0010067750699818134
step: 30, loss: 0.00038567313458770514
step: 40, loss: 0.03111613728106022
step: 50, loss: 0.0038409288972616196
step: 60, loss: 0.0004218601970933378
step: 70, loss: 0.03477807343006134
step: 80, loss: 0.0006016494007781148
step: 90, loss: 0.0792384222149849
step: 100, loss: 0.0006231782026588917
step: 110, loss: 0.0007488814881071448
step: 120, loss: 0.0054376511834561825
step: 130, loss: 0.00793190486729145
step: 140, loss: 0.021544067189097404
step: 150, loss: 0.0031765836756676435
step: 160, loss: 0.05803897976875305
step: 170, loss: 0.0023926706053316593
step: 180, loss: 0.003332659602165222
step: 190, loss: 0.024227211251854897
step: 200, loss: 0.00020216350094415247
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 210, loss: 0.00015841939602978528
step: 220, loss: 0.0005941493436694145
step: 230, loss: 0.0002417588693788275
step: 240, loss: 0.14609554409980774
step: 250, loss: 0.0002974431263282895
step: 260, loss: 0.021068433299660683
step: 270, loss: 0.03663903474807739
step: 280, loss: 0.0965530276298523
step: 290, loss: 0.0006743383128196001
step: 300, loss: 0.0028830477967858315
step: 310, loss: 0.0005420593661256135
step: 320, loss: 0.004182380624115467
step: 330, loss: 0.0004137785581406206
step: 340, loss: 0.0017521497793495655
step: 350, loss: 0.005657467059791088
step: 360, loss: 0.07841019332408905
step: 370, loss: 0.0005548147018998861
step: 380, loss: 0.006499126087874174
step: 390, loss: 0.004310602322220802
step: 400, loss: 0.0007591368048451841
step: 410, loss: 0.0004362605686765164
step: 420, loss: 0.003918330650776625
```

```
step: 430, loss: 0.0029251310043036938
step: 440, loss: 0.01835322193801403
step: 450, loss: 0.004841144196689129
step: 460, loss: 0.0004263095906935632
step: 470, loss: 0.016626812517642975
step: 480, loss: 0.0024496987462043762
step: 490, loss: 0.023668449372053146
step: 500, loss: 0.022548088803887367
step: 510, loss: 0.0002971129142679274
step: 520, loss: 0.00014012088649906218
step: 530, loss: 0.0030482790898531675
/usr/local/lib/python3.7/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=0.9431192660550459, f1=0.9410681399631675,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002154189714929089
step: 10, loss: 0.017647143453359604
step: 20, loss: 0.0009761160472407937
step: 30, loss: 0.0008577814442105591
step: 40, loss: 0.0003121408517472446
step: 50, loss: 0.10040437430143356
step: 60, loss: 0.0046175140887498856
step: 70, loss: 0.005893330089747906
step: 80, loss: 0.0010727026965469122
step: 90, loss: 0.00044304304174147546
step: 100, loss: 0.04135838896036148
step: 110, loss: 0.0011045344872400165
step: 120, loss: 0.00035204592859372497
step: 130, loss: 0.004392137285321951
step: 140, loss: 0.00015820871340110898
step: 150, loss: 0.005541244521737099
step: 160, loss: 0.0002544415183365345
step: 170, loss: 6.93631372996606e-05
step: 180, loss: 9.884985047392547e-05
step: 190, loss: 0.011829276569187641
step: 200, loss: 0.007745067123323679
step: 210, loss: 7.342931348830462e-05
step: 220, loss: 0.001291286083869636
step: 230, loss: 0.013438858091831207
step: 240, loss: 0.0033444901928305626
step: 250, loss: 0.014442534185945988
step: 260, loss: 0.1304236799478531
step: 270, loss: 0.00018109929806087166
step: 280, loss: 0.0007720545982010663
```

```
step: 290, loss: 0.0013742925366386771
step: 300, loss: 0.00029042395181022584
step: 310, loss: 0.0803440734744072
step: 320, loss: 0.009212767705321312
step: 330, loss: 0.00039477780228480697
step: 340, loss: 0.005320493597537279
step: 350, loss: 0.0609886571764946
step: 360, loss: 0.0005389350117184222
step: 370, loss: 0.0028327456675469875
step: 380, loss: 0.0015329087618738413
step: 390, loss: 0.0016894323052838445
step: 400, loss: 6.383912113960832e-05
step: 410, loss: 0.00011021128739230335
step: 420, loss: 0.07824841886758804
step: 430, loss: 0.000686021929141134
step: 440, loss: 0.0031360811553895473
step: 450, loss: 0.0010688183829188347
step: 460, loss: 0.07415871322154999
step: 470, loss: 0.05851864442229271
step: 480, loss: 0.009069683961570263
step: 490, loss: 0.00015756633365526795
step: 500, loss: 0.012589900754392147
step: 510, loss: 0.00021039061539340764
step: 520, loss: 0.004360325634479523
step: 530, loss: 0.0010003320639953017
/usr/local/lib/python3.7/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=0.9405756731662025, f1=0.9355140186915888,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013217394007369876
step: 10, loss: 0.0022974114399403334
step: 20, loss: 0.01048656739294529
step: 30, loss: 0.0034569697454571724
step: 40, loss: 0.000481313414638862
step: 50, loss: 0.0018803321290761232
step: 60, loss: 0.0021753953769803047
step: 70, loss: 0.0913827195763588
step: 80, loss: 0.06037793681025505
step: 90, loss: 0.009731349535286427
step: 100, loss: 0.00019649631576612592
step: 110, loss: 0.0007128600846044719
step: 120, loss: 0.00011768731201300398
step: 130, loss: 0.00032646485487930477
step: 140, loss: 0.0014399834908545017
```

```
step: 150, loss: 4.076766344951466e-05
step: 160, loss: 8.637477003503591e-05
step: 170, loss: 0.0005969351623207331
step: 180, loss: 5.542063809116371e-05
step: 190, loss: 0.00021075000404380262
step: 200, loss: 0.0004735394904855639
step: 210, loss: 0.005474085919559002
step: 220, loss: 0.00014212897804100066
step: 230, loss: 0.008661152794957161
step: 240, loss: 0.008837857283651829
step: 250, loss: 0.00041653955122455955
step: 260, loss: 0.0027306056581437588
step: 270, loss: 0.00033792920294217765
step: 280, loss: 0.042238879948854446
step: 290, loss: 0.0029308209195733307
step: 300, loss: 0.00019267783500254154
step: 310, loss: 0.000784377392847091
step: 320, loss: 0.027513224631547928
step: 330, loss: 6.844892050139606e-05
step: 340, loss: 0.011561441235244274
step: 350, loss: 0.00018806647858582437
step: 360, loss: 0.0013596613425761461
step: 370, loss: 7.798171282047406e-05
step: 380, loss: 9.317092917626724e-05
step: 390, loss: 3.959755122195929e-05
step: 400, loss: 4.7542544052703306e-05
step: 410, loss: 0.00014099576219450682
step: 420, loss: 9.325609426014125e-05
step: 430, loss: 5.1302817155374214e-05
step: 440, loss: 3.564141661627218e-05
step: 450, loss: 0.07396920025348663
step: 460, loss: 0.0010522211669012904
step: 470, loss: 0.004019368439912796
step: 480, loss: 0.001210157759487629
step: 490, loss: 7.330565131269395e-05
step: 500, loss: 5.860976307303645e-05
step: 510, loss: 0.0018139773746952415
step: 520, loss: 0.00014913876657374203
step: 530, loss: 4.811254621017724e-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=0.9417889256980597, f1=0.9351763584366063,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015892315423116088
```

```
step: 10, loss: 0.005315955728292465
step: 20, loss: 0.00876785907894373
step: 30, loss: 0.017421728000044823
step: 40, loss: 6.66995401843451e-05
step: 50, loss: 0.0029522059485316277
step: 60, loss: 4.45659679826349e-05
step: 70, loss: 6.248088175198063e-05
step: 80, loss: 6.249739817576483e-05
step: 90, loss: 6.010862853145227e-05
step: 100, loss: 0.0003679541405290365
step: 110, loss: 9.469747601542622e-05
step: 120, loss: 0.004255807958543301
step: 130, loss: 0.0005051622283644974
step: 140, loss: 3.657991692307405e-05
step: 150, loss: 0.00013193876657169312
step: 160, loss: 4.8772817535791546e-05
step: 170, loss: 6.658276106463745e-05
step: 180, loss: 0.00013354919792618603
step: 190, loss: 5.0753144023474306e-05
step: 200, loss: 0.0005107845645397902
step: 210, loss: 0.00021608083625324073
step: 220, loss: 0.0005808654823340476
step: 230, loss: 0.00010088845010614023
step: 240, loss: 5.463070556288585e-05
step: 250, loss: 0.05662324279546738
step: 260, loss: 0.005655941087752581
step: 270, loss: 6.385925371432677e-05
step: 280, loss: 0.0003559703764040023
step: 290, loss: 0.00015329477901104838
step: 300, loss: 0.0002849758311640471
step: 310, loss: 0.007634412497282028
step: 320, loss: 7.306014595087618e-05
step: 330, loss: 0.0027658892795443535
step: 340, loss: 4.588550655171275e-05
step: 350, loss: 0.005640360992401838
step: 360, loss: 0.0001769968803273514
step: 370, loss: 0.00020726259390357882
step: 380, loss: 0.00864359550178051
step: 390, loss: 0.00034269780735485256
step: 400, loss: 0.13436609506607056
step: 410, loss: 0.015107057988643646
step: 420, loss: 6.87336505507119e-05
step: 430, loss: 0.0005294909351505339
step: 440, loss: 0.00032937867217697203
step: 450, loss: 0.00024986988864839077
step: 460, loss: 0.0009450636571273208
step: 470, loss: 0.00014525304140988737
step: 480, loss: 0.0003689067088998854
```

```
step: 490, loss: 0.0002473767090123147
step: 500, loss: 0.0006534553831443191
step: 510, loss: 0.0006332965567708015
step: 520, loss: 0.0003193924203515053
step: 530, loss: 0.0004793744010385126
/usr/local/lib/python3.7/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=0.9434315100514259, f1=0.9392523364485983,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003970940597355366
step: 10, loss: 0.00029248339706100523
step: 20, loss: 0.03864113986492157
step: 30, loss: 0.00011815250036306679
step: 40, loss: 0.005102287046611309
step: 50, loss: 0.00021114967239554971
step: 60, loss: 8.65935580804944e-05
step: 70, loss: 0.05581015348434448
step: 80, loss: 0.0002865197020582855
step: 90, loss: 0.015598123893141747
step: 100, loss: 0.0001247143663931638
step: 110, loss: 0.000643695006147027
step: 120, loss: 6.729361484758556e-05
step: 130, loss: 7.150353485485539e-05
step: 140, loss: 0.007294507697224617
step: 150, loss: 0.0001857443858170882
step: 160, loss: 0.0058335126377642155
step: 170, loss: 3.873800233122893e-05
step: 180, loss: 0.00019540467474143952
step: 190, loss: 0.0017836211482062936
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 200, loss: 0.00011192984675290063
step: 210, loss: 4.1091872844845057e-05
step: 220, loss: 0.001524782506749034
step: 230, loss: 0.0007415224681608379
step: 240, loss: 8.001111564226449e-05
step: 250, loss: 0.00017434007895644754
step: 260, loss: 0.001512475311756134
step: 270, loss: 0.0010686656460165977
step: 280, loss: 4.8337318730773404e-05
step: 290, loss: 3.8558977394131944e-05
step: 300, loss: 7.122172246454284e-05
step: 310, loss: 6.168879917822778e-05
step: 320, loss: 0.0003586771781556308
step: 330, loss: 0.004673885181546211
```

```
step: 340, loss: 0.00021515753178391606
step: 350, loss: 0.0002194705157307908
step: 360, loss: 0.018759911879897118
step: 370, loss: 0.00010029767145169899
step: 380, loss: 3.596274837036617e-05
step: 390, loss: 0.00010463173384778202
step: 400, loss: 0.016668986529111862
step: 410, loss: 0.00011121848365291953
step: 420, loss: 0.0005461085238493979
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 430, loss: 4.159841773798689e-05
step: 440, loss: 0.009176366031169891
step: 450, loss: 3.382458817213774e-05
step: 460, loss: 0.00013313638919498771
step: 470, loss: 0.0022222811821848154
step: 480, loss: 0.0007738444255664945
step: 490, loss: 4.4388336391421035e-05
step: 500, loss: 0.013727198354899883
step: 510, loss: 0.00016932805010583252
step: 520, loss: 0.00013397129077930003
step: 530, loss: 5.117234104545787e-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=0.9448818897637795, f1=0.9441860465116279,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012069364078342915
step: 10, loss: 7.481482316507027e-05
step: 20, loss: 6.163014040794224e-05
step: 30, loss: 7.616260700160637e-05
step: 40, loss: 0.00012955439160577953
step: 50, loss: 0.0010496203321963549
step: 60, loss: 0.0071311271749436855
step: 70, loss: 0.00021782907424494624
step: 80, loss: 0.00018330689636059105
step: 90, loss: 0.0003126043011434376
step: 100, loss: 0.00015425127639900893
step: 110, loss: 0.0015309526352211833
step: 120, loss: 9.922243771143258e-05
step: 130, loss: 0.005500964354723692
step: 140, loss: 7.115831249393523e-05
step: 150, loss: 4.461126809474081e-05
step: 160, loss: 0.00025931576965376735
step: 170, loss: 6.751824548700824e-05
step: 180, loss: 0.00010630913311615586
```

```
step: 190, loss: 0.00012140359467593953
step: 200, loss: 0.00022873950365465134
step: 210, loss: 8.08986442280002e-05
step: 220, loss: 0.0001491146394982934
step: 230, loss: 0.014069166034460068
step: 240, loss: 4.1352250264026225e-05
step: 250, loss: 5.309242260409519e-05
step: 260, loss: 0.022622739896178246
step: 270, loss: 0.00034117669565603137
step: 280, loss: 4.1768947994569317e-05
step: 290, loss: 6.267917342483997e-05
step: 300, loss: 0.0212709978222847
step: 310, loss: 5.0537288188934326e-05
step: 320, loss: 3.011002809216734e-05
step: 330, loss: 3.555638613761403e-05
step: 340, loss: 3.81927820853889e-05
step: 350, loss: 4.280866414774209e-05
step: 360, loss: 0.00010126063716597855
step: 370, loss: 8.451915346086025e-05
step: 380, loss: 3.2800191547721624e-05
step: 390, loss: 3.3578577131265774e-05
step: 400, loss: 0.00012096003774786368
step: 410, loss: 0.004764128942042589
step: 420, loss: 0.00015176544548012316
step: 430, loss: 9.89191685221158e-05
step: 440, loss: 0.001372727332636714
step: 450, loss: 0.0006313982885330915
step: 460, loss: 0.0009651202708482742
step: 470, loss: 6.946251232875511e-05
step: 480, loss: 7.426854426739737e-05
step: 490, loss: 0.027369746938347816
step: 500, loss: 0.011238967068493366
step: 510, loss: 9.700059308670461e-05
step: 520, loss: 0.00011255324352532625
step: 530, loss: 3.779912003665231e-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=0.9447282861124013, f1=0.9373246024321795,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00019111853907816112
step: 10, loss: 0.010864811949431896
step: 20, loss: 7.941509102238342e-05
step: 30, loss: 0.0005041955737397075
step: 40, loss: 0.0018431661883369088
```

```
step: 50, loss: 0.00013547579874284565
step: 60, loss: 5.9455098380567506e-05
step: 70, loss: 0.0005355006433092058
step: 80, loss: 0.00010905876115430146
step: 90, loss: 0.0015834571095183492
step: 100, loss: 3.3552718377904966e-05
step: 110, loss: 8.681495819473639e-05
step: 120, loss: 0.00028879655292257667
step: 130, loss: 3.708720760187134e-05
step: 140, loss: 4.361841274658218e-05
step: 150, loss: 0.00016977843188215047
step: 160, loss: 0.00010314321843907237
step: 170, loss: 0.0003543501952663064
step: 180, loss: 6.072331962059252e-05
step: 190, loss: 0.00024798666709110269
step: 200, loss: 0.017188360914587975
step: 210, loss: 3.811940405284986e-05
step: 220, loss: 5.554005110752769e-05
step: 230, loss: 0.00011520410771481693
step: 240, loss: 3.641678631538525e-05
step: 250, loss: 0.0006049683433957398
step: 260, loss: 3.363044743309729e-05
step: 270, loss: 5.649848753819242e-05
step: 280, loss: 7.215841469587758e-05
step: 290, loss: 0.000601058651227504
step: 300, loss: 0.028038669377565384
step: 310, loss: 9.45531646721065e-05
step: 320, loss: 0.018215902149677277
step: 330, loss: 3.893142275046557e-05
step: 340, loss: 6.098105222918093e-05
step: 350, loss: 0.0003089249657932669
step: 360, loss: 4.955687836627476e-05
step: 370, loss: 2.143498022633139e-05
step: 380, loss: 0.03768043965101242
step: 390, loss: 0.0012751995818689466
step: 400, loss: 3.003200254170224e-05
step: 410, loss: 2.6177065592492e-05
step: 420, loss: 0.18861356377601624
step: 430, loss: 0.0003195374447386712
step: 440, loss: 4.822875780519098e-05
step: 450, loss: 6.582825153600425e-05
step: 460, loss: 0.003578250063583255
step: 470, loss: 0.007438452448695898
step: 480, loss: 2.9704713597311638e-05
step: 490, loss: 4.831553451367654e-05
step: 500, loss: 8.25772003736347e-05
step: 510, loss: 3.081486647715792e-05
step: 520, loss: 2.541334652050864e-05
```

```
step: 530, loss: 3.0103108656476252e-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=0.9447565543071161, f1=0.941286989196806,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00048930844059214
step: 10, loss: 0.018283667042851448
step: 20, loss: 2.2775837351218797e-05
step: 30, loss: 6.437707634177059e-05
step: 40, loss: 0.0001895671884994954
step: 50, loss: 0.008384189568459988
step: 60, loss: 3.6930501664755866e-05
step: 70, loss: 0.003314228495582938
step: 80, loss: 0.0001468545524403453
step: 90, loss: 0.04636850953102112
step: 100, loss: 2.4660814233357087e-05
step: 110, loss: 5.098054316476919e-05
step: 120, loss: 9.107726509682834e-05
step: 130, loss: 8.979152335086837e-05
step: 140, loss: 0.00037792144576087594
step: 150, loss: 0.08748131990432739
step: 160, loss: 0.00023803692602086812
step: 170, loss: 0.00031921896152198315
step: 180, loss: 7.57980378693901e-05
step: 190, loss: 3.1101379136089236e-05
step: 200, loss: 6.926427886355668e-05
step: 210, loss: 8.512673230143264e-05
step: 220, loss: 0.0002501298440620303
step: 230, loss: 0.00014143054431769997
step: 240, loss: 4.417386662680656e-05
step: 250, loss: 2.737263093877118e-05
step: 260, loss: 0.00010877299791900441
step: 270, loss: 3.1775027309777215e-05
step: 280, loss: 2.8247995942365378e-05
step: 290, loss: 0.00398745434358716
step: 300, loss: 7.90084304753691e-05
step: 310, loss: 4.156792419962585e-05
step: 320, loss: 0.00013006810331717134
step: 330, loss: 2.3297470761463046e-05
step: 340, loss: 3.8415397284552455e-05
step: 350, loss: 0.02849072404205799
step: 360, loss: 4.274621096556075e-05
step: 370, loss: 2.4630946427350864e-05
step: 380, loss: 4.1239381971536204e-05
```

```
step: 390, loss: 3.0270384741015732e-05
step: 400, loss: 3.97885414713528e-05
step: 410, loss: 0.0001731950615067035
step: 420, loss: 4.594373604049906e-05
step: 430, loss: 4.2647843656595796e-05
step: 440, loss: 7.496993202948943e-05
step: 450, loss: 0.0011681203031912446
step: 460, loss: 0.0006391496281139553
step: 470, loss: 3.561380071914755e-05
step: 480, loss: 4.0769529732642695e-05
step: 490, loss: 7.068749255267903e-05
step: 500, loss: 0.00021699469652958214
step: 510, loss: 9.65252474998124e-05
step: 520, loss: 0.023624230176210403
step: 530, loss: 0.00010906332317972556
/usr/local/lib/python3.7/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=0.9454545454545454, f1=0.936768149882904,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010480867786100134
step: 10, loss: 1.815680843719747e-05
step: 20, loss: 0.023546310141682625
step: 30, loss: 3.8152844354044646e-05
step: 40, loss: 2.9688922950299457e-05
step: 50, loss: 2.914192918979097e-05
step: 60, loss: 2.951071110146586e-05
step: 70, loss: 2.534234408813063e-05
step: 80, loss: 2.3710930690867826e-05
step: 90, loss: 4.729912325274199e-05
step: 100, loss: 1.8633563740877435e-05
step: 110, loss: 0.0004338233848102391
step: 120, loss: 0.0001215875381603837
step: 130, loss: 2.7722389859263785e-05
step: 140, loss: 1.818285818444565e-05
step: 150, loss: 0.0002683925849851221
step: 160, loss: 2.1431182176456787e-05
step: 170, loss: 2.7219470212003216e-05
step: 180, loss: 2.2920990886632353e-05
step: 190, loss: 3.806237145909108e-05
step: 200, loss: 6.547741941176355e-05
step: 210, loss: 4.3294312490615994e-05
step: 220, loss: 2.6437252017785795e-05
step: 230, loss: 9.568730456521735e-05
step: 240, loss: 2.1744155674241483e-05
```

```
step: 250, loss: 0.029507186263799667
step: 260, loss: 2.7926222173846327e-05
step: 270, loss: 0.00039745704270899296
step: 280, loss: 1.8480528524378315e-05
step: 290, loss: 3.9327329432126135e-05
step: 300, loss: 8.332495053764433e-05
step: 310, loss: 0.00015004743181634694
step: 320, loss: 2.0336025045253336e-05
step: 330, loss: 0.0015813139034435153
step: 340, loss: 4.031842036056332e-05
step: 350, loss: 0.06451746076345444
step: 360, loss: 3.1762843718752265e-05
step: 370, loss: 3.4685774153331295e-05
step: 380, loss: 0.0008301148773171008
step: 390, loss: 3.440536238485947e-05
step: 400, loss: 0.00011270985123701394
step: 410, loss: 8.441402314929292e-05
step: 420, loss: 0.0001550493761897087
step: 430, loss: 0.002920733764767647
step: 440, loss: 0.00012121615145588294
step: 450, loss: 5.8175803133053705e-05
step: 460, loss: 2.7158966986462474e-05
step: 470, loss: 0.026011060923337936
step: 480, loss: 0.000659306941088289
step: 490, loss: 0.00011443971743574366
step: 500, loss: 0.000490523234475404
step: 510, loss: 2.1025078240199946e-05
step: 520, loss: 2.0384291929076426e-05
step: 530, loss: 2.7677622711053118e-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=0.9432029795158287, f1=0.9373831775700935,
best_f1=0.9343955014058105
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.0434475522488356e-05
step: 10, loss: 8.809519931674004e-05
step: 20, loss: 3.0038703698664904e-05
step: 30, loss: 2.6698460715124384e-05
step: 40, loss: 0.00011099460971308872
step: 50, loss: 3.0475406674668193e-05
step: 60, loss: 2.9878661734983325e-05
step: 70, loss: 5.2952240366721526e-05
step: 80, loss: 7.886737148510292e-05
step: 90, loss: 1.4051626749278512e-05
step: 100, loss: 9.97542156255804e-05
```

```
step: 110, loss: 2.249614044558257e-05
step: 120, loss: 2.5207496946677566e-05
step: 130, loss: 2.7219011826673523e-05
step: 140, loss: 0.0002438308292767033
step: 150, loss: 0.00012297822104301304
step: 160, loss: 0.0007418552413582802
step: 170, loss: 0.00029887325945310295
step: 180, loss: 3.4500892070354894e-05
step: 190, loss: 4.0623937820782885e-05
step: 200, loss: 7.122836541384459e-05
step: 210, loss: 7.418382301693782e-05
step: 220, loss: 8.544621960027143e-05
step: 230, loss: 0.002215554006397724
step: 240, loss: 5.975887688691728e-05
step: 250, loss: 6.86695784679614e-05
step: 260, loss: 1.4923338312655687e-05
step: 270, loss: 0.001114986022002995
step: 280, loss: 0.00015769402671139687
step: 290, loss: 2.5334604288218543e-05
step: 300, loss: 3.35836703015957e-05
step: 310, loss: 1.6905090888030827e-05
step: 320, loss: 0.00014937296509742737
step: 330, loss: 2.8257729354663752e-05
step: 340, loss: 3.3317577617708594e-05
step: 350, loss: 0.00013885952648706734
step: 360, loss: 0.002203586744144559
step: 370, loss: 0.00026544462889432907
step: 380, loss: 1.693120429990813e-05
step: 390, loss: 4.228638135828078e-05
step: 400, loss: 2.2816930140834302e-05
step: 410, loss: 2.249604403914418e-05
step: 420, loss: 6.154397851787508e-05
step: 430, loss: 2.229524943686556e-05
step: 440, loss: 1.1831426490971353e-05
step: 450, loss: 3.195651152054779e-05
step: 460, loss: 6.29774367553182e-05
step: 470, loss: 1.599987081135623e-05
step: 480, loss: 2.2768157577957027e-05
step: 490, loss: 2.69923621090012902e-05
step: 500, loss: 0.00011922104749828577
step: 510, loss: 4.58402355434373e-05
step: 520, loss: 3.0941253498895094e-05
step: 530, loss: 2.30553159781266e-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=0.9440820130475303, f1=0.9418386491557222,
best_f1=0.9343955014058105
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.584118010418024e-05
step: 10, loss: 4.1399860492674634e-05
step: 20, loss: 2.01981147256447e-05
step: 30, loss: 4.3558262404985726e-05
step: 40, loss: 2.193641739722807e-05
step: 50, loss: 0.0006928701186552644
step: 60, loss: 1.2833486835006624e-05
step: 70, loss: 3.614290108089335e-05
step: 80, loss: 2.166585181839764e-05
step: 90, loss: 1.6130192307173274e-05
step: 100, loss: 4.702664591604844e-05
step: 110, loss: 0.01591872051358223
step: 120, loss: 0.00037622664240188897
step: 130, loss: 1.3470502381096594e-05
step: 140, loss: 0.0013665154110640287
step: 150, loss: 6.092103649280034e-05
step: 160, loss: 4.272500882507302e-05
step: 170, loss: 2.019416933762841e-05
step: 180, loss: 2.175434201490134e-05
step: 190, loss: 0.0007338121067732573
step: 200, loss: 2.0887151549686678e-05
step: 210, loss: 2.8194719561724924e-05
step: 220, loss: 1.2017679182463326e-05
step: 230, loss: 0.00020563842554111034
step: 240, loss: 5.484030771185644e-05
step: 250, loss: 5.301575947669335e-05
step: 260, loss: 4.066008114023134e-05
step: 270, loss: 0.0001745784393278882
step: 280, loss: 0.00017619389109313488
step: 290, loss: 0.00013068862608633935
step: 300, loss: 0.0018214347073808312
step: 310, loss: 3.8946131098782644e-05
step: 320, loss: 2.9055994673399255e-05
step: 330, loss: 4.456897659110837e-05
step: 340, loss: 4.385550710139796e-05
step: 350, loss: 0.00013677187962457538
step: 360, loss: 0.0012193639995530248
step: 370, loss: 1.718818020890467e-05
step: 380, loss: 2.9817301765433513e-05
step: 390, loss: 2.9916804123786278e-05
step: 400, loss: 1.507975048298249e-05
step: 410, loss: 0.0001215140800923109
step: 420, loss: 3.888506762450561e-05
step: 430, loss: 2.9800865377183072e-05
step: 440, loss: 0.00047681829892098904
```

```
step: 450, loss: 0.001289958250708878
step: 460, loss: 1.6074323866632767e-05
step: 470, loss: 1.581723518029321e-05
step: 480, loss: 2.2414169507101178e-05
step: 490, loss: 0.00020962893904652447
step: 500, loss: 0.0024860987905412912
step: 510, loss: 3.475142875686288e-05
step: 520, loss: 3.0522700399160385e-05
step: 530, loss: 1.8257300325785764e-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=0.9439252336448598, f1=0.939679547596607,
best_f1=0.9343955014058105
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:16, 355.35it/s]
load_f1 = 0.9475620975160993
real_f1 = 0.9462563160312356
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, 355.51it/s]
```

3 DK

```
[8]:  # Para executar 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 99.7 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: 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: wasabi<1.1.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (0.10.1)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (1.0.0)
Requirement already satisfied: 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: 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: 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: 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: 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: 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: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.12.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5)
(3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5)
(4.1.1)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5)
(2.10)
Requirement already satisfied: 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: 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: 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=94c3946d65adf8057307a842697d0f499cbd36f703a98b88319ae0dd8effe7af
  Stored in directory: /tmp/pip-ephem-wheel-cache-
rijf9caa/wheels/11/95/ba/2c36cc368c0bd339b44a791c2c1881a1fb714b78c29a4cb8f5
Successfully built en-core-web-lg
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-2.2.5
 Download and installation successful
You can now load the model via spacy.load('en_core_web_lg')

3.1 DK STRUCTURED

3.1.1 Beer - Running the matcher

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

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

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

Selected optimization level 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)…

245

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: 0.8531317710876465
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.27184466019417475, f1=0.2666666666666667,
best_f1=0.2666666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37622275948524475
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.2947368421052632, f1=0.288659793814433,
best_f1=0.288659793814433
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.34272313117980957
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.45283018867924535, f1=0.40740740740740744,
best_f1=0.40740740740740744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.35706713795661926
/usr/local/lib/python3.7/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=0.45, f1=0.35555555555555557, best_f1=0.40740740740740744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.26761046051979065
/usr/local/lib/python3.7/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=0.4482758620689656, f1=0.4210526315789474,
best_f1=0.40740740740740744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.26695045828819275
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5185185185185185, f1=0.3728813559322034,
best_f1=0.3728813559322034
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1883220225572586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.5, f1=0.4, best_f1=0.3728813559322034
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3224965035915375
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5714285714285714, f1=0.391304347826087,
best_f1=0.391304347826087
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15225449204444885
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4615384615384615,
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: 0.23046700656414032
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6666666666666665, f1=0.3902439024390244,
best_f1=0.3902439024390244
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2317926585674286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6875000000000001, f1=0.3888888888888889,
best_f1=0.3888888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24656783044338226
/usr/local/lib/python3.7/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=0.6451612903225806, f1=0.3888888888888889,
best_f1=0.3888888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13269901275634766
/usr/local/lib/python3.7/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=0.6428571428571429, f1=0.3870967741935484,
best_f1=0.3888888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15771181881427765
/usr/local/lib/python3.7/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=0.6451612903225806, f1=0.4, best_f1=0.3888888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2676737308502197
/usr/local/lib/python3.7/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=0.6451612903225806, f1=0.4, best_f1=0.3888888888888889
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.")
91it [00:00, 136950.72it/s]
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
load_f1 = 0.6153846153846153
real_f1 = 0.5714285714285714
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 [00:09, 466.52it/s]
```

3.1.2 DBLP-ACM - Running the matcher

```python
# 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 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: 0.8084027171134949
step: 10, loss: 0.4710726737976074
step: 20, loss: 0.5734274983406067
step: 30, loss: 0.376738041639328
step: 40, loss: 0.12557052075862885
step: 50, loss: 0.03582189604640007
step: 60, loss: 0.05580252408981323
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.22235094010829926
step: 80, loss: 0.08122368156909943
step: 90, loss: 0.007757021114230156
step: 100, loss: 0.06151115149259567
step: 110, loss: 0.026454854756593704
step: 120, loss: 0.019165033474564552
step: 130, loss: 0.006706941872835159
step: 140, loss: 0.005194745026528835
step: 150, loss: 0.07334037870168686
step: 160, loss: 0.08871693164110184
step: 170, loss: 0.031344834715127945
step: 180, loss: 0.010345309972763062
step: 190, loss: 0.011843468993902206
step: 200, loss: 0.002467183396220207
step: 210, loss: 0.005702031776309013
step: 220, loss: 0.007886883802711964
step: 230, loss: 0.009778440929949284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9853107344632768, 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: 0.054474301636219025
step: 10, loss: 0.0020712995901703835
step: 20, loss: 0.007124720141291618
step: 30, loss: 0.0030746501870453358

```
step: 40, loss: 0.0037562232464551926
step: 50, loss: 0.011292294599115849
step: 60, loss: 0.0030793205369263887
step: 70, loss: 0.05369871109724045
step: 80, loss: 0.00489934254437685
step: 90, loss: 0.003979983739554882
step: 100, loss: 0.0218567606061697
step: 110, loss: 0.030358312651515007
step: 120, loss: 0.0037841936573386192
step: 130, loss: 0.0008315143059007823
step: 140, loss: 0.024154121056199074
step: 150, loss: 0.008464816957712173
step: 160, loss: 0.0058410074561834335
step: 170, loss: 0.02212970331311226
step: 180, loss: 0.002588442759588361
step: 190, loss: 0.027882009744644165
step: 200, loss: 0.0011196924606338143
step: 210, loss: 0.10804572701454163
step: 220, loss: 0.0014617822598665953
step: 230, loss: 0.03388063609600067
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.990990990990991, f1=0.9865168539325843,
best_f1=0.9865168539325843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13071812689304352
step: 10, loss: 0.0014306738739833236
step: 20, loss: 0.00783497467637062
step: 30, loss: 0.024578450247645378
step: 40, loss: 0.023283658549189568
step: 50, loss: 0.014578862115740776
step: 60, loss: 0.0011613041860982776
step: 70, loss: 0.0010210119653493166
step: 80, loss: 0.09268394857645035
step: 90, loss: 0.01113755814731121
step: 100, loss: 0.0017699033487588167
step: 110, loss: 0.0050337729044258595
step: 120, loss: 0.005714279133826494
step: 130, loss: 0.0010093628661707044
step: 140, loss: 0.006304511800408363
step: 150, loss: 0.001374053186737001
```

```
step: 160, loss: 0.0015092482790350914
step: 170, loss: 0.004784406162798405
step: 180, loss: 0.001438905019313097
step: 190, loss: 0.0018748854054138064
step: 200, loss: 0.01790865696966648
step: 210, loss: 0.024512091651558876
step: 220, loss: 0.001077030086889863
step: 230, loss: 0.003009392414242029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9932279909706545, 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: 0.00048443191917761315
step: 10, loss: 0.0008360301144421101
step: 20, loss: 0.00024200460757128894
step: 30, loss: 0.00038470583967864513
step: 40, loss: 0.0011665355414152145
step: 50, loss: 0.0067074974067509174
step: 60, loss: 0.0011823789682239294
step: 70, loss: 0.06611350923776627
step: 80, loss: 0.09978778660297394
step: 90, loss: 0.007341054268181324
step: 100, loss: 0.0017913259798660874
step: 110, loss: 0.007173565682023764
step: 120, loss: 0.013029446825385094
step: 130, loss: 0.0005903732962906361
step: 140, loss: 0.0004635776858776808
step: 150, loss: 0.0005741639761254191
step: 160, loss: 0.0013159450609236956
step: 170, loss: 0.006383034400641918
step: 180, loss: 0.14257977902889252
step: 190, loss: 0.011868222616612911
step: 200, loss: 0.0010092725278809667
step: 210, loss: 0.0011208057403564453
step: 220, loss: 0.0008497939561493695
step: 230, loss: 0.0003910030936822295
/usr/local/lib/python3.7/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=0.9920903954802259, f1=0.9853768278965129,
```

```
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: 0.00048521667486056685
step: 10, loss: 0.0004262711154296994
step: 20, loss: 0.00017963176651392132
step: 30, loss: 0.00011633505346253514
step: 40, loss: 0.0004384742060210556
step: 50, loss: 0.03672536090016365
step: 60, loss: 0.0003549082321114838
step: 70, loss: 0.00019676881493069232
step: 80, loss: 0.0007038459298200905
step: 90, loss: 0.00199719937518239
step: 100, loss: 0.003759734332561493
step: 110, loss: 0.0033319424837827682
step: 120, loss: 0.07905695587396622
step: 130, loss: 0.11426443606615067
step: 140, loss: 0.0007183421985246241
step: 150, loss: 0.00021773797925561666
step: 160, loss: 0.09969106316566467
step: 170, loss: 0.019237572327256203
step: 180, loss: 0.0018337105866521597
step: 190, loss: 0.0012344641145318747
step: 200, loss: 0.009410830214619637
step: 210, loss: 0.0002971524663735181
step: 220, loss: 0.000394139759009704
step: 230, loss: 0.0006198401679284871
/usr/local/lib/python3.7/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=0.9909706546275394, f1=0.987598647125141,
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: 0.0007365899509750307
step: 10, loss: 0.00048300609341822565
step: 20, loss: 0.001628332887776196
step: 30, loss: 0.0024450018536299467
step: 40, loss: 0.0010165160056203604
step: 50, loss: 0.0033435628283768892
step: 60, loss: 0.04778997227549553
step: 70, loss: 0.0011887059081345797
step: 80, loss: 0.004527818877249956
step: 90, loss: 0.00038300661253742874
step: 100, loss: 0.0007369755767285824
step: 110, loss: 0.05496717989444733
```

```
step: 120, loss: 0.0005317740142345428
step: 130, loss: 0.04045776650309563
step: 140, loss: 0.00151852669660002
step: 150, loss: 0.0054925004951655865
step: 160, loss: 0.05082295835018158
step: 170, loss: 0.0013200129615142941
step: 180, loss: 0.005920444615185261
step: 190, loss: 0.02882932685315609
step: 200, loss: 0.0011927911546081305
step: 210, loss: 0.0007471583085134625
step: 220, loss: 0.0005606170743703842
step: 230, loss: 0.0013739196583628654
/usr/local/lib/python3.7/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=0.9876819708846584, f1=0.9832026875699889,
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: 0.06443998217582703
step: 10, loss: 0.0004231375933159143
step: 20, loss: 0.0004949294379912317
step: 30, loss: 0.0007529936265200377
step: 40, loss: 0.0019403300248086452
step: 50, loss: 0.00011948934843530878
step: 60, loss: 0.005932385101914406
step: 70, loss: 0.000246876705205068
step: 80, loss: 0.00012463676102925092
step: 90, loss: 0.0003408488992135972
step: 100, loss: 0.0007076008478179574
step: 110, loss: 0.032889388501644135
step: 120, loss: 0.0006633390439674258
step: 130, loss: 0.00026130900369025767
step: 140, loss: 0.00015552766853943467
step: 150, loss: 0.0003599321353249252
step: 160, loss: 0.00017746408411767334
step: 170, loss: 0.0001484045060351491
step: 180, loss: 0.001676173647865653
step: 190, loss: 0.005954025778919458
step: 200, loss: 0.0007029006956145167
step: 210, loss: 0.00010472907888470218
step: 220, loss: 0.00034449330996721983
step: 230, loss: 0.07985705882310867
/usr/local/lib/python3.7/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=0.9829738933030647, f1=0.9852774631936579,
```

```
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: 0.09075872600078583
step: 10, loss: 0.04119110852479935
step: 20, loss: 0.0004207503516227007
step: 30, loss: 0.00016957157640717924
step: 40, loss: 0.0029540054965764284
step: 50, loss: 0.001767751295119524
step: 60, loss: 0.00043127761455252767
step: 70, loss: 0.0002165222686016932
step: 80, loss: 0.0014342541107907891
step: 90, loss: 0.00012331221166013345
step: 100, loss: 0.00014700704196002334
step: 110, loss: 0.0001638131943764165
step: 120, loss: 6.473559915320948e-05
step: 130, loss: 0.00042911438504233956
step: 140, loss: 8.084187720669433e-05
step: 150, loss: 0.00017743541684467345
step: 160, loss: 0.00014998999540694058
step: 170, loss: 0.0005557402619160712
step: 180, loss: 0.00018181692576035857
step: 190, loss: 0.0001836221053963527
step: 200, loss: 0.00034628447610884905
step: 210, loss: 6.690014561172575e-05
step: 220, loss: 7.589269080199301e-05
step: 230, loss: 0.03592148423194885
/usr/local/lib/python3.7/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=0.9876543209876544, f1=0.9832026875699889,
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: 0.0005676518194377422
step: 10, loss: 8.634563710074872e-05
step: 20, loss: 0.00048635012353770435
step: 30, loss: 0.0003041280433535576
step: 40, loss: 4.673051444115117e-05
step: 50, loss: 0.00013438657333608717
step: 60, loss: 0.00022676325170323253
step: 70, loss: 0.00021606851078104228
step: 80, loss: 0.05259395390748978
step: 90, loss: 0.0009811894269660115
step: 100, loss: 0.0004292545490898192
step: 110, loss: 0.00019157763745170087
```

```
step: 120, loss: 0.028160328045487404
step: 130, loss: 9.034344111569226e-05
step: 140, loss: 9.310481254942715e-05
step: 150, loss: 6.316081999102607e-05
step: 160, loss: 8.846202399581671e-05
step: 170, loss: 5.213898839429021e-05
step: 180, loss: 8.786393300397322e-05
step: 190, loss: 4.881345375906676e-05
step: 200, loss: 8.778005576459691e-05
step: 210, loss: 8.115964737953618e-05
step: 220, loss: 0.040762919932603836
step: 230, loss: 0.00242908578366041
/usr/local/lib/python3.7/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=0.9819819819819819, 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: 7.50139806768857e-05
step: 10, loss: 8.455548959318548e-05
step: 20, loss: 0.00012361377594061196
step: 30, loss: 8.597533451393247e-05
step: 40, loss: 9.563787898514424e-05
step: 50, loss: 0.0014617005363106728
step: 60, loss: 0.018152153119444847
step: 70, loss: 9.493909601587802e-05
step: 80, loss: 0.0028747469186782837
step: 90, loss: 8.513174543622881e-05
step: 100, loss: 0.00012599093315657228
step: 110, loss: 0.0003594241279643029
step: 120, loss: 0.014331265352666378
step: 130, loss: 6.265551201067865e-05
step: 140, loss: 0.004825639072805643
step: 150, loss: 0.0006465267506428063
step: 160, loss: 0.0001637401874177158
step: 170, loss: 8.012235048227012e-05
step: 180, loss: 0.0006095542339608073
step: 190, loss: 0.00020850343571510166
step: 200, loss: 7.889309199526906e-05
step: 210, loss: 0.00012487189087551087
step: 220, loss: 0.0014312757411971688
step: 230, loss: 0.00012085682101314887
/usr/local/lib/python3.7/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=0.9853768278965129, 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: 7.07429280737415e-05
step: 10, loss: 0.0003092041879426688
step: 20, loss: 4.302791421650909e-05
step: 30, loss: 0.00014045368880033493
step: 40, loss: 0.00013282123836688697
step: 50, loss: 0.00023781333584338427
step: 60, loss: 8.29646160127595e-05
step: 70, loss: 0.00018914291285909712
step: 80, loss: 8.967050962382928e-05
step: 90, loss: 5.409821460489184e-05
step: 100, loss: 6.99690353940241e-05
step: 110, loss: 0.00024782525724731386
step: 120, loss: 9.569120447849855e-05
step: 130, loss: 0.00014444654516410083
step: 140, loss: 0.0013311079237610102
step: 150, loss: 4.4819873437518254e-05
step: 160, loss: 0.0028938439209014177
step: 170, loss: 0.010901708155870438
step: 180, loss: 5.4948130127741024e-05
step: 190, loss: 8.561831782571971e-05
step: 200, loss: 5.246436921879649e-05
step: 210, loss: 4.745649130200036e-05
step: 220, loss: 4.437343159224838e-05
step: 230, loss: 0.018645694479346275
/usr/local/lib/python3.7/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=0.9795454545454545, f1=0.976271186440678,
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: 0.00032598551479168236
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.004532896913588047
step: 20, loss: 5.2381117711775005e-05
step: 30, loss: 0.0005183861358091235
step: 40, loss: 6.114347343100235e-05
step: 50, loss: 0.00011178421118529513
step: 60, loss: 0.010247541591525078
step: 70, loss: 0.00011019128578482196
step: 80, loss: 8.575033280067146e-05
step: 90, loss: 4.1042389057110995e-05
step: 100, loss: 6.632657459704205e-05
```

```
step: 110, loss: 7.863883365644142e-05
step: 120, loss: 6.005119576002471e-05
step: 130, loss: 6.42825907561928e-05
step: 140, loss: 9.04864864423871e-05
step: 150, loss: 4.27152517659124e-05
step: 160, loss: 0.026149500161409378
step: 170, loss: 0.00010791842214530334
step: 180, loss: 9.410237544216216e-05
step: 190, loss: 0.00030789032462053
step: 200, loss: 0.0037254062481224537
step: 210, loss: 4.714758324553259e-05
step: 220, loss: 0.029715510085225105
step: 230, loss: 4.8013109335443005e-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=0.9898074745186863, f1=0.9887133182844244,
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: 0.00014466288848780096
step: 10, loss: 7.388601079583168e-05
step: 20, loss: 8.680723840370774e-05
step: 30, loss: 0.03410976752638817
step: 40, loss: 0.0004711965739261359
step: 50, loss: 4.791239189216867e-05
step: 60, loss: 4.852638812735677e-05
step: 70, loss: 4.096538759768009e-05
step: 80, loss: 4.7937868657754734e-05
step: 90, loss: 3.6316687328508124e-05
step: 100, loss: 6.168145773699507e-05
step: 110, loss: 4.444825026439503e-05
step: 120, loss: 0.00012227213301230222
step: 130, loss: 6.929822848178446e-05
step: 140, loss: 0.00013854741700924933
step: 150, loss: 0.017998451367020607
step: 160, loss: 3.095625652349554e-05
step: 170, loss: 5.24444694747217e-05
step: 180, loss: 6.011482764733955e-05
step: 190, loss: 3.2073479815153405e-05
step: 200, loss: 0.00010316038242308423
step: 210, loss: 6.332305201794952e-05
step: 220, loss: 3.7119250919204205e-05
step: 230, loss: 3.773900607484393e-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=0.9887133182844244, f1=0.9831649831649831,
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: 3.636111432570033e-05
step: 10, loss: 2.8266478693694808e-05
step: 20, loss: 0.011850851587951183
step: 30, loss: 0.00010519889474380761
step: 40, loss: 3.247180939069949e-05
step: 50, loss: 4.7427300160052255e-05
step: 60, loss: 3.6129902582615614e-05
step: 70, loss: 4.446997627383098e-05
step: 80, loss: 0.00011178404383827001
step: 90, loss: 0.00020372758444864303
step: 100, loss: 0.0031565355602651834
step: 110, loss: 8.209225052269176e-05
step: 120, loss: 3.4722273994702846e-05
step: 130, loss: 7.737474516034126e-05
step: 140, loss: 0.0001585610443726182
step: 150, loss: 6.913490506121889e-05
step: 160, loss: 5.256346412352286e-05
step: 170, loss: 3.6338649806566536e-05
step: 180, loss: 6.730500899720937e-05
step: 190, loss: 0.0025518066249787807
step: 200, loss: 3.694233237183653e-05
step: 210, loss: 0.009697610512375832
step: 220, loss: 6.322939589153975e-05
step: 230, loss: 2.94696928904159e-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=0.9866071428571428,
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: 3.9280206692637876e-05
step: 10, loss: 4.482427902985364e-05
step: 20, loss: 4.68015277874656e-05
step: 30, loss: 0.000281641841866076
step: 40, loss: 8.857793000061065e-05
step: 50, loss: 6.65655461489223e-05
step: 60, loss: 3.797743920586072e-05
step: 70, loss: 4.417998206918128e-05
step: 80, loss: 0.02442394755780697
step: 90, loss: 2.819971450662706e-05
step: 100, loss: 7.301868754439056e-05

```
step: 110, loss: 6.341046537272632e-05
step: 120, loss: 7.682070281589404e-05
step: 130, loss: 5.8224162785336375e-05
step: 140, loss: 6.508918158942834e-05
step: 150, loss: 0.00038468887214548886
step: 160, loss: 4.88344048790168e-05
step: 170, loss: 3.458785795373842e-05
step: 180, loss: 7.734746759524569e-05
step: 190, loss: 0.004170096479356289
step: 200, loss: 4.811346298083663e-05
step: 210, loss: 0.028682414442300797
step: 220, loss: 0.00023392782895825803
step: 230, loss: 6.572740676347166e-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=0.9909297052154196, f1=0.987598647125141,
best_f1=0.9887640449438202
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
2473it [00:07, 350.37it/s]
load_f1 = 0.990990990990991
real_f1 = 0.9898534385569334
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 423.06it/s]
```

3.1.3 DBLP-GoogleScholar - Running the matcher

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


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

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

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

Defaults for this optimization level are:
```

263

```
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: 0.7910195589065552
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4400377869606018
step: 20, loss: 0.48721855878829956
step: 30, loss: 0.41642534732818604
step: 40, loss: 0.2642582952976227
step: 50, loss: 0.16498056054115295
step: 60, loss: 0.14531677961349487
step: 70, loss: 0.2397921085357666
step: 80, loss: 0.08064180612564087
step: 90, loss: 0.10866688936948776
step: 100, loss: 0.2664280831813812
step: 110, loss: 0.09263446182012558
step: 120, loss: 0.04916013404726982
step: 130, loss: 0.025398431345820427
step: 140, loss: 0.28210219740867615
```

```
step: 150, loss: 0.06351442635059357
step: 160, loss: 0.11760223656892776
step: 170, loss: 0.09240804612636566
step: 180, loss: 0.10416232794523239
step: 190, loss: 0.050950221717357635
step: 200, loss: 0.0835372805595398
step: 210, loss: 0.07752252370119095
step: 220, loss: 0.07564574480056763
step: 230, loss: 0.11816014349460602
step: 240, loss: 0.08541461080312729
step: 250, loss: 0.06998157501220703
step: 260, loss: 0.023890197277069092
step: 270, loss: 0.025038881227374077
step: 280, loss: 0.11116073280572891
step: 290, loss: 0.10610942542552948
step: 300, loss: 0.12501755356788635
step: 310, loss: 0.08430448174476624
step: 320, loss: 0.03939303010702133
step: 330, loss: 0.060564469546079636
step: 340, loss: 0.18596181273460388
step: 350, loss: 0.07301350682973862
step: 360, loss: 0.0722585916519165
step: 370, loss: 0.15627625584602356
step: 380, loss: 0.19950179755687714
step: 390, loss: 0.04533160850405693
step: 400, loss: 0.010021556168794632
step: 410, loss: 0.011992176063358784
step: 420, loss: 0.005888128187507391
step: 430, loss: 0.04239843413233757
step: 440, loss: 0.05537993088364601
step: 450, loss: 0.009029863402247429
step: 460, loss: 0.12872669100761414
step: 470, loss: 0.06749482452869415
step: 480, loss: 0.3024584650993347
step: 490, loss: 0.0278138667345047
step: 500, loss: 0.02113707736134529
step: 510, loss: 0.09162832796573639
step: 520, loss: 0.02442205511033535
step: 530, loss: 0.08338964730501175
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9318826868495743, f1=0.9297094657919399,
best_f1=0.9297094657919399
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11214135587215424
step: 10, loss: 0.12349250167608261
step: 20, loss: 0.10809151828289032
step: 30, loss: 0.030214082449674606
step: 40, loss: 0.017356716096401215
step: 50, loss: 0.08662697672843933
step: 60, loss: 0.12281102687120438
step: 70, loss: 0.13383394479751587
step: 80, loss: 0.013723577372729778
step: 90, loss: 0.004884790163487196
step: 100, loss: 0.3064369559288025
step: 110, loss: 0.010211176238954067
step: 120, loss: 0.07508572190999985
step: 130, loss: 0.0660046860575676
step: 140, loss: 0.040752772241830826
step: 150, loss: 0.0359322614967823
step: 160, loss: 0.04329773038625717
step: 170, loss: 0.12586738169193268
step: 180, loss: 0.01877412013709545
step: 190, loss: 0.01269037276506424
step: 200, loss: 0.005669753532856703
step: 210, loss: 0.00436184648424387
step: 220, loss: 0.14825639128684998
step: 230, loss: 0.020444057881832123
step: 240, loss: 0.10086666792631149
step: 250, loss: 0.017810460180044174
step: 260, loss: 0.007988939061760902
step: 270, loss: 0.22147543728351593
step: 280, loss: 0.2640427052974701
step: 290, loss: 0.09300138056278229
step: 300, loss: 0.04218735173344612
step: 310, loss: 0.04717443883419037
step: 320, loss: 0.08649107813835144
step: 330, loss: 0.04710241034626961
step: 340, loss: 0.034123752266168594
step: 350, loss: 0.017029203474521637
step: 360, loss: 0.03197913616895676
step: 370, loss: 0.01098252646625042
step: 380, loss: 0.12443815916776657
step: 390, loss: 0.025827931240200996
step: 400, loss: 0.0637284517288208
step: 410, loss: 0.0024581430479884148
step: 420, loss: 0.039537250995635986
step: 430, loss: 0.035184651613235474
step: 440, loss: 0.022448046132922173
```

```
step: 450, loss: 0.019345907494425774
step: 460, loss: 0.15274463593959808
step: 470, loss: 0.03163040429353714
step: 480, loss: 0.260288804769516
step: 490, loss: 0.03741467371582985
step: 500, loss: 0.016940202564001083
step: 510, loss: 0.06542187184095383
step: 520, loss: 0.209908202229053497
step: 530, loss: 0.1373732089996338
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9464944649446495, f1=0.9425815342214057,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004760828334838152
step: 10, loss: 0.029188774526119232
step: 20, loss: 0.09307622909545898
step: 30, loss: 0.26798978447914124
step: 40, loss: 0.0062188031151890755
step: 50, loss: 0.005258017219603062
step: 60, loss: 0.009265953674912453
step: 70, loss: 0.07281970232725143
step: 80, loss: 0.0023519464302808046
step: 90, loss: 0.036923352628946304
step: 100, loss: 0.0176073405891656888
step: 110, loss: 0.006301144603639841
step: 120, loss: 0.020843282341957092
step: 130, loss: 0.07529327273368835
step: 140, loss: 0.07485876232385635
step: 150, loss: 0.006183736026287079
step: 160, loss: 0.002636995166540146
step: 170, loss: 0.0021588944364339113
step: 180, loss: 0.017454974353313446
step: 190, loss: 0.0011717365123331547
step: 200, loss: 0.0033178848680108786
step: 210, loss: 0.06043524295091629
step: 220, loss: 0.00661378912627697
step: 230, loss: 0.010957974009215832
step: 240, loss: 0.004020656459033489
step: 250, loss: 0.021286990493535995
step: 260, loss: 0.0009987286757677794
```

```
step: 270, loss: 0.005566976964473724
step: 280, loss: 0.0066322158090770245
step: 290, loss: 0.02685249224305153
step: 300, loss: 0.02591131441295147
step: 310, loss: 0.05506928265094757
step: 320, loss: 0.1189911961555481
step: 330, loss: 0.002054058713838458
step: 340, loss: 0.0011460771784186363
step: 350, loss: 0.07013498991727829
step: 360, loss: 0.07768232375383377
step: 370, loss: 0.0009372307686135173
step: 380, loss: 0.07527169585227966
step: 390, loss: 0.10120297968387604
step: 400, loss: 0.029566776007413864
step: 410, loss: 0.012935122475028038
step: 420, loss: 0.022667333483695984
step: 430, loss: 0.05143951252102852
step: 440, loss: 0.016815047711133957
step: 450, loss: 0.09194277971982956
step: 460, loss: 0.0892704427242279
step: 470, loss: 0.0033446757588535547
step: 480, loss: 0.01902601681649685
step: 490, loss: 0.004537599161267281
step: 500, loss: 0.026729736477136612
step: 510, loss: 0.004076474811881781
step: 520, loss: 0.0014581069117411971
step: 530, loss: 0.01199844665825367
/usr/local/lib/python3.7/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=0.9422180801491147, f1=0.9433085501858736,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032444042153656483
step: 10, loss: 0.010324565693736076
step: 20, loss: 0.03409157320857048
step: 30, loss: 0.01601264625787735
step: 40, loss: 0.02890860289335251
step: 50, loss: 0.0017131430795416236
step: 60, loss: 0.0005542519502341747
step: 70, loss: 0.0043877712450921535
step: 80, loss: 0.07025039196014404
step: 90, loss: 0.0016261255368590355
step: 100, loss: 0.05285749211907387
step: 110, loss: 0.01079445518553257
step: 120, loss: 0.0005260962061583996
```

```
step: 130, loss: 0.04658304899930954
step: 140, loss: 0.014489840716123581
step: 150, loss: 0.014163381420075893
step: 160, loss: 0.014954978600144386
step: 170, loss: 0.0029628644697368145
step: 180, loss: 0.0014383280649781227
step: 190, loss: 0.0702255517244339
step: 200, loss: 0.003376817563548684
step: 210, loss: 0.037184685468673706
step: 220, loss: 0.0018282720120623708
step: 230, loss: 0.17882075905799866
step: 240, loss: 0.004547027871012688
step: 250, loss: 0.0001839999749790877
step: 260, loss: 0.09758684039115906
step: 270, loss: 0.025169963017106056
step: 280, loss: 0.0019984908867627382
step: 290, loss: 0.16649723052978516
step: 300, loss: 0.002638396807014942
step: 310, loss: 0.006216009147465229
step: 320, loss: 0.008615529164671898
step: 330, loss: 0.059727273881435394
step: 340, loss: 0.0019917343743145466
step: 350, loss: 0.0011315104784443974
step: 360, loss: 0.009162291884422302
step: 370, loss: 0.039827700704336166
step: 380, loss: 0.0015389099717140198
step: 390, loss: 0.01822168566286564
step: 400, loss: 0.004805325996130705
step: 410, loss: 0.03755146265029907
step: 420, loss: 0.008717340417206287
step: 430, loss: 0.007980276830494404
step: 440, loss: 0.06826279312372208
step: 450, loss: 0.00661282567307353
step: 460, loss: 0.008103689178824425
step: 470, loss: 0.010622200556099415
step: 480, loss: 0.0007498597260564566
step: 490, loss: 0.049810197204351425
step: 500, loss: 0.005148690659552813
step: 510, loss: 0.006618532817810774
step: 520, loss: 0.0063783046789467335
step: 530, loss: 0.0101195452734828
/usr/local/lib/python3.7/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=0.9391304347826086, f1=0.9380127620783956,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0022046726662665606
step: 10, loss: 0.0015387285966426134
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 0.00128695007879287
step: 30, loss: 0.0008113168296404183
step: 40, loss: 0.039485324174165726
step: 50, loss: 0.03826681524515152
step: 60, loss: 0.0017468651058152318
step: 70, loss: 0.00026621436700224876
step: 80, loss: 0.0036423890851438046
step: 90, loss: 0.000713743269443512
step: 100, loss: 0.000212754457606934
step: 110, loss: 0.0034246183931827545
step: 120, loss: 0.004021159373223782
step: 130, loss: 0.0026987497694790363
step: 140, loss: 0.009475599974393845
step: 150, loss: 0.0016736150719225407
step: 160, loss: 0.0011754364240914583
step: 170, loss: 0.008246338926255703
step: 180, loss: 0.0017270274693146348
step: 190, loss: 0.0001349341036984697
step: 200, loss: 0.25062376260757446
step: 210, loss: 0.0009203705121763051
step: 220, loss: 0.006916572339832783
step: 230, loss: 0.0028724109288305044
step: 240, loss: 0.002354080555960536
step: 250, loss: 0.004700460471212864
step: 260, loss: 0.0004270944918971509
step: 270, loss: 0.0022944968659430742
step: 280, loss: 0.04277815669775009
step: 290, loss: 0.00025524638476781547
step: 300, loss: 0.0059295217506587505
step: 310, loss: 0.0008860501693561673
step: 320, loss: 0.004791381768882275
step: 330, loss: 0.0006183767691254616
step: 340, loss: 0.0015719863586127758
step: 350, loss: 0.025952065363526344
step: 360, loss: 0.00445935083553195
step: 370, loss: 0.025855451822280884
step: 380, loss: 0.03468228876590729
step: 390, loss: 0.0017755954759195447
step: 400, loss: 0.0744595155119896
step: 410, loss: 0.0004019752377644181
step: 420, loss: 0.008502746932208538
step: 430, loss: 0.023930439725518227
step: 440, loss: 0.0020564559381455183
step: 450, loss: 0.0007088023703545332
```

```
step: 460, loss: 0.0001912919688038528
step: 470, loss: 0.0037247100844979286
step: 480, loss: 0.015686452388763428
step: 490, loss: 0.0028643300756812096
step: 500, loss: 0.0010718912817537785
step: 510, loss: 0.011063911952078342
step: 520, loss: 0.00039001123514026403
step: 530, loss: 0.010720551945269108
/usr/local/lib/python3.7/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=0.9423791821561338, f1=0.9358736059479555,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012138254242017865
step: 10, loss: 0.0013265479356050491
step: 20, loss: 0.012849793769419193
step: 30, loss: 0.01686088740825653
step: 40, loss: 0.0012611736310645938
step: 50, loss: 0.034931376576423645
step: 60, loss: 0.00048460892867296934
step: 70, loss: 0.05304737761616707
step: 80, loss: 0.0006313227931968868
step: 90, loss: 0.0052407970651984215
step: 100, loss: 0.0001594458444742486
step: 110, loss: 0.0012045209296047688
step: 120, loss: 0.00018917694978881627
step: 130, loss: 0.006192626431584358
step: 140, loss: 0.00011052543413825333
step: 150, loss: 0.0020451245363801718
step: 160, loss: 8.321792120113969e-05
step: 170, loss: 0.007105973083525896
step: 180, loss: 0.00027883017901331186
step: 190, loss: 0.001371570280753076
step: 200, loss: 0.00012795663496945053
step: 210, loss: 0.0005273143178783357
step: 220, loss: 0.001964221941307187
step: 230, loss: 0.00011830370203824714
step: 240, loss: 9.313542977906764e-05
step: 250, loss: 6.5076797909569e-05
step: 260, loss: 0.0026293352711945772
step: 270, loss: 0.001095083192922175
step: 280, loss: 0.0011191709199920297
step: 290, loss: 0.006397772580385208
step: 300, loss: 0.0034667765721678734
step: 310, loss: 0.000864157045725733
```

```
step: 320, loss: 0.0078712757676683983
step: 330, loss: 0.004994216840714216
step: 340, loss: 0.03015468455851078
step: 350, loss: 0.08085706830024719
step: 360, loss: 0.001597945811226964
step: 370, loss: 0.0018199196783825755
step: 380, loss: 0.08858148753643036
step: 390, loss: 0.012268955819308758
step: 400, loss: 0.00026814505690708756
step: 410, loss: 0.006062173284590244
step: 420, loss: 0.050300389528274536
step: 430, loss: 0.03826364502310753
step: 440, loss: 0.04418249800801277
step: 450, loss: 0.0011171492515131831
step: 460, loss: 0.000554168364033103
step: 470, loss: 0.06072656437754631
step: 480, loss: 0.009982890449464321
step: 490, loss: 0.004504729062318802
step: 500, loss: 0.00563350273296237
step: 510, loss: 0.00041594341746531427
step: 520, loss: 0.009900900535285473
step: 530, loss: 0.0001314718829235062
/usr/local/lib/python3.7/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=0.9410672853828307, f1=0.9431345353675451,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014413699973374605
step: 10, loss: 0.0071894102272985697
step: 20, loss: 0.0027020371053367853
step: 30, loss: 0.0006557214655913413
step: 40, loss: 8.483911369694397e-05
step: 50, loss: 0.00046378260594792664
step: 60, loss: 0.054214902222156525
step: 70, loss: 0.001037876820191741
step: 80, loss: 0.0004797782457899302
step: 90, loss: 0.00011488945165183395
step: 100, loss: 9.959291492123157e-05
step: 110, loss: 0.0002738609036896378
step: 120, loss: 0.025565652176737785
step: 130, loss: 9.122995106736198e-05
step: 140, loss: 0.00016015941218938679
step: 150, loss: 4.8999805585481226e-05
step: 160, loss: 0.0007614321657456458
step: 170, loss: 0.00018345507851336151
```

```
step: 180, loss: 0.0001548567379359156
step: 190, loss: 4.7629382606828585e-05
step: 200, loss: 0.00010627945448504761
step: 210, loss: 0.0006817773100920022
step: 220, loss: 0.0006178191979415715
step: 230, loss: 6.885327456984669e-05
step: 240, loss: 0.01206224225461483
step: 250, loss: 0.00012600506306625903
step: 260, loss: 6.744256825186312e-05
step: 270, loss: 4.2060335545102134e-05
step: 280, loss: 0.0009454928804188967
step: 290, loss: 0.0017268465599045157
step: 300, loss: 0.00021866355382371694
step: 310, loss: 0.0003031733212992549
step: 320, loss: 0.016542349010705948
step: 330, loss: 0.00036217522574588656
step: 340, loss: 0.0009466076153330505
step: 350, loss: 0.005264188162982464
step: 360, loss: 0.004169354680925608
step: 370, loss: 0.0019764502067118883
step: 380, loss: 0.013037439435720444
step: 390, loss: 0.0002730327541939914
step: 400, loss: 0.0001328717771684751
step: 410, loss: 0.0020461599342525005
step: 420, loss: 0.0010839373571798205
step: 430, loss: 0.026776306331157684
step: 440, loss: 9.851681534200907e-05
step: 450, loss: 0.10850004851818085
step: 460, loss: 0.00140280963387340
step: 470, loss: 0.016895104199647903
step: 480, loss: 0.005709092132747173
step: 490, loss: 0.0006799369584769011
step: 500, loss: 0.0001629549660719931
step: 510, loss: 0.0006347293383441865
step: 520, loss: 0.00018642212671693414
step: 530, loss: 4.862052810494788e-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=0.9419953596287702, f1=0.9423165666820489,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004284601891413331
step: 10, loss: 0.01621071621775627
step: 20, loss: 0.022039027884602547
step: 30, loss: 0.008805489167571068
```

```
step: 40, loss: 0.02814515493810177
step: 50, loss: 0.00016596910427324474
step: 60, loss: 0.0015791059704497457
step: 70, loss: 0.00037546863313764334
step: 80, loss: 0.03885626047849655
step: 90, loss: 0.06668823957443237
step: 100, loss: 0.0008065203437581658
step: 110, loss: 0.0005551102804020047
step: 120, loss: 0.02562306448817253
step: 130, loss: 0.05067940056324005
step: 140, loss: 3.436125552980229e-05
step: 150, loss: 0.0023934992495924234
step: 160, loss: 0.0009855515090748668
step: 170, loss: 0.07150594145059586
step: 180, loss: 0.00022547232219949365
step: 190, loss: 9.120622416958213e-05
step: 200, loss: 0.0009829680202528834
step: 210, loss: 0.0046824305318295956
step: 220, loss: 0.0008364488021470606
step: 230, loss: 0.0018574268324300647
step: 240, loss: 0.00023123690334614366
step: 250, loss: 0.0012058715801686049
step: 260, loss: 0.016285153105854988
step: 270, loss: 3.321754411444999e-05
step: 280, loss: 0.00010473634756635875
step: 290, loss: 0.00011100734991487116
step: 300, loss: 0.0021368134766817093
step: 310, loss: 0.04826191067695618
step: 320, loss: 8.10824494692497e-05
step: 330, loss: 0.03807668015360832
step: 340, loss: 0.0004820131289307028
step: 350, loss: 0.00015436504327226043
step: 360, loss: 0.0004254973609931767
step: 370, loss: 0.0001347556826658547
step: 380, loss: 0.0021219071932137012
step: 390, loss: 0.00013370325905270875
step: 400, loss: 0.015242747962474823
step: 410, loss: 0.0005999410641379654
step: 420, loss: 3.823791121249087e-05
step: 430, loss: 0.0018860184354707599
step: 440, loss: 0.0003908336511813104
step: 450, loss: 0.00032714573899284005
step: 460, loss: 0.002116207731887698
step: 470, loss: 0.00010949412535410374
step: 480, loss: 0.0014576244866475463
step: 490, loss: 3.628679769462906e-05
step: 500, loss: 4.284627357264981e-05
step: 510, loss: 2.7678424885380082e-05
```

```
step: 520, loss: 0.0006757648661732674
step: 530, loss: 4.785452620126307e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9384328358208955, f1=0.9413936317489617,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.199078052304685e-05
step: 10, loss: 4.806054130312987e-05
step: 20, loss: 3.8968370063230395e-05
step: 30, loss: 6.887236668262631e-05
step: 40, loss: 0.004286837298423052
step: 50, loss: 0.0001132784818764776
step: 60, loss: 8.273391722468659e-05
step: 70, loss: 0.10167151689529419
step: 80, loss: 0.000268906878773123
step: 90, loss: 9.07511857803911e-05
step: 100, loss: 0.0007410714752040803
step: 110, loss: 0.003954144194722176
step: 120, loss: 5.870727909496054e-05
step: 130, loss: 0.008901509456336498
step: 140, loss: 0.0015466585755348206
step: 150, loss: 4.272209480404854e-05
step: 160, loss: 4.4685999455396086e-05
step: 170, loss: 3.499057493172586e-05
step: 180, loss: 0.0004139643278904259
step: 190, loss: 0.0001260960998479277
step: 200, loss: 5.8761750551639125e-05
step: 210, loss: 3.69866902474314e-05
step: 220, loss: 5.100753696751781e-05
step: 230, loss: 3.824930536211468e-05
step: 240, loss: 4.873011857853271e-05
step: 250, loss: 0.00013137406494934112
step: 260, loss: 0.0001530535373603925
step: 270, loss: 0.00047897628974169493
step: 280, loss: 4.231690400047228e-05
step: 290, loss: 2.7648520699585788e-05
step: 300, loss: 0.0005578970885835588
step: 310, loss: 2.7532894819159992e-05
step: 320, loss: 0.00013205295545049012
step: 330, loss: 0.0014385540271177888
step: 340, loss: 0.01060451939702034
step: 350, loss: 0.020895186811685562
step: 360, loss: 0.015846027061343193
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
```

```
step: 370, loss: 0.004264642484486103
step: 380, loss: 0.0002591885859146714
step: 390, loss: 0.0003185477980878204
step: 400, loss: 0.005533203948289156
step: 410, loss: 0.0003070073144044727
step: 420, loss: 0.0010993358446285129
step: 430, loss: 4.286882540327497e-05
step: 440, loss: 0.0001290041400352493
step: 450, loss: 0.15559427440166473
step: 460, loss: 0.00024382858828175813
step: 470, loss: 0.00865256693214178
step: 480, loss: 0.04324442520737648
step: 490, loss: 0.0004142938705626875
step: 500, loss: 0.008289019577205181
step: 510, loss: 0.00023570602934341878
step: 520, loss: 0.0001833548885770142
step: 530, loss: 0.0015806688461452723
/usr/local/lib/python3.7/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=0.9459962756052142, f1=0.9445471349353051,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001030646963045001
step: 10, loss: 7.365465717157349e-05
step: 20, loss: 0.004573758225888014
step: 30, loss: 3.969492536270991e-05
step: 40, loss: 0.0007456045132130384
step: 50, loss: 5.132190926815383e-05
step: 60, loss: 0.00014182183076627553
step: 70, loss: 8.653007535031065e-05
step: 80, loss: 0.0014936523512005806
step: 90, loss: 0.00023408916604239494
step: 100, loss: 0.002205059165135026
step: 110, loss: 0.0010736532276496291
step: 120, loss: 0.0007916682516224682
step: 130, loss: 0.00012652113218791783
step: 140, loss: 0.0002013530465774238
step: 150, loss: 5.1809918659273535e-05
step: 160, loss: 0.00034517922904342413
step: 170, loss: 0.0003502311883494258
step: 180, loss: 0.0030550716910511255
step: 190, loss: 0.00010947415285045281
step: 200, loss: 7.724217721261084e-05
step: 210, loss: 0.03728127107024193
step: 220, loss: 0.0003803927975241095
```

```
step: 230, loss: 0.0005644371267408133
step: 240, loss: 5.949178375885822e-05
step: 250, loss: 0.009761331602931023
step: 260, loss: 0.014868387952446938
step: 270, loss: 0.029031939804553986
step: 280, loss: 5.097800385556184e-05
step: 290, loss: 4.2594932892825454e-05
step: 300, loss: 0.00011201352026546374
step: 310, loss: 0.000496993656270206
step: 320, loss: 0.0007601312245242298
step: 330, loss: 0.000336984870955348
step: 340, loss: 0.00011027412256225944
step: 350, loss: 0.00037457101279869676
step: 360, loss: 8.210411033360288e-05
step: 370, loss: 0.0021502641029655933
step: 380, loss: 6.0823971580248326e-05
step: 390, loss: 5.3614356147591025e-05
step: 400, loss: 0.0003514106210786849
step: 410, loss: 0.003078989451751113
step: 420, loss: 0.00016728616901673377
step: 430, loss: 0.00023980512924026698
step: 440, loss: 5.330407657311298e-05
step: 450, loss: 0.003105779178440571
step: 460, loss: 0.008492942899465561
step: 470, loss: 2.7652222343022004e-05
step: 480, loss: 0.0008836158085614443
step: 490, loss: 0.01621835120022297
step: 500, loss: 0.0015428701881319284
step: 510, loss: 0.0001599319657543674
step: 520, loss: 0.00021346623543649912
step: 530, loss: 0.055634159594774246
/usr/local/lib/python3.7/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=0.9409576940957695, f1=0.9444958371877892,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003029039129961483
step: 10, loss: 0.0025575861800462008
step: 20, loss: 0.00042277725879102945
step: 30, loss: 0.004817393608391285
step: 40, loss: 5.52214914932847e-05
step: 50, loss: 9.141518967226148e-05
step: 60, loss: 0.0005517348181456327
step: 70, loss: 2.8579348509083502e-05
step: 80, loss: 0.0001115105624194257
```

```
step: 90, loss: 0.03226562961935997
step: 100, loss: 3.184588058502413e-05
step: 110, loss: 0.0016420641914010048
step: 120, loss: 0.0001288395724259317
step: 130, loss: 2.1464720703079365e-05
step: 140, loss: 2.0246672647772357e-05
step: 150, loss: 0.0003167229879181832
step: 160, loss: 5.4116328101372346e-05
step: 170, loss: 0.0007164202979765832
step: 180, loss: 0.0005173814715817571
step: 190, loss: 0.00011333519796608016
step: 200, loss: 0.0004119396617170423
step: 210, loss: 3.1012128602014855e-05
step: 220, loss: 3.199950515409e-05
step: 230, loss: 0.00011623774480540305
step: 240, loss: 2.3487527869292535e-05
step: 250, loss: 7.035073213046417e-05
step: 260, loss: 3.686033232952468e-05
step: 270, loss: 0.008210557512938976
step: 280, loss: 0.20460684597492218
step: 290, loss: 0.0036605915520340204
step: 300, loss: 0.0025776068214327097
step: 310, loss: 0.005147472023963928
step: 320, loss: 0.02023053914308548
step: 330, loss: 0.0018282983219251037
step: 340, loss: 0.0001609960017958656
step: 350, loss: 0.0002995660179294646
step: 360, loss: 9.284477710025385e-05
step: 370, loss: 1.991513454413507e-05
step: 380, loss: 3.3011830964824185e-05
step: 390, loss: 0.0010424703359603882
step: 400, loss: 3.066007775487378e-05
step: 410, loss: 4.604465721058659e-05
step: 420, loss: 0.0012309425510466099
step: 430, loss: 0.0002902511914726347
step: 440, loss: 2.7901829525944777e-05
step: 450, loss: 2.7264079108135775e-05
step: 460, loss: 0.0005982521688565612
step: 470, loss: 0.0009428717312403023
step: 480, loss: 0.00031058292370289564
step: 490, loss: 0.000548557611182332
step: 500, loss: 3.3451546187279746e-05
step: 510, loss: 0.06941195577383041
step: 520, loss: 2.5618104700697586e-05
step: 530, loss: 2.278701686009299e-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=0.9418874941887495, f1=0.9411764705882353,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006944636697880924
step: 10, loss: 3.137715975753963e-05
step: 20, loss: 1.66965546668850716e-05
step: 30, loss: 3.815981108346023e-05
step: 40, loss: 3.678048597066663e-05
step: 50, loss: 8.794834138825536e-05
step: 60, loss: 0.00024318424402736127
step: 70, loss: 0.0011397635098546743
step: 80, loss: 0.0018909493228420615
step: 90, loss: 6.560319161508232e-05
step: 100, loss: 0.013509661890566349
step: 110, loss: 2.4239823687821627e-05
step: 120, loss: 0.00012428985792212188
step: 130, loss: 2.963758925034199e-05
step: 140, loss: 2.0790534108527936e-05
step: 150, loss: 3.8285754271782935e-05
step: 160, loss: 3.802221908699721e-05
step: 170, loss: 2.7367579605197534e-05
step: 180, loss: 4.45815094280988e-05
step: 190, loss: 0.006843827664852142
step: 200, loss: 0.0012799418764188886
step: 210, loss: 7.531569281127304e-05
step: 220, loss: 2.1058705897303298e-05
step: 230, loss: 0.0021259894128888845
step: 240, loss: 9.752339974511415e-05
step: 250, loss: 2.7342623070580885e-05
step: 260, loss: 4.568382792058401e-05
step: 270, loss: 3.403975279070437e-05
step: 280, loss: 9.074155968846753e-05
step: 290, loss: 0.0005417817155830562
step: 300, loss: 0.00019743957091122866
step: 310, loss: 3.394229133846238e-05
step: 320, loss: 2.7275713364360854e-05
step: 330, loss: 3.879520590999164e-05
step: 340, loss: 0.006781686097383499
step: 350, loss: 0.017578478902578354
step: 360, loss: 0.001292042201384902
step: 370, loss: 1.836543196986895e-05
step: 380, loss: 0.1428035944700241
step: 390, loss: 2.2507476387545466e-05
step: 400, loss: 2.0008212231914513e-05
step: 410, loss: 0.0097526116296648984
step: 420, loss: 0.0012691693846136332
```

```
step: 430, loss: 4.7641802666475996e-05
step: 440, loss: 6.992195267230272e-05
step: 450, loss: 6.389205373125151e-05
step: 460, loss: 0.0002769277198240161
step: 470, loss: 2.026518814091105e-05
step: 480, loss: 5.240141763351858e-05
step: 490, loss: 0.0008392469026148319
step: 500, loss: 0.005184745416045189
step: 510, loss: 3.944845593650825e-05
step: 520, loss: 0.051721639931201935
step: 530, loss: 0.001930141355842352
/usr/local/lib/python3.7/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=0.9330783938814532, f1=0.9373814041745729,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003626411489676684
step: 10, loss: 0.00010843555355677381
step: 20, loss: 0.020564598962664604
step: 30, loss: 2.8415643100743182e-05
step: 40, loss: 9.336009679827839e-05
step: 50, loss: 2.0913370462949388e-05
step: 60, loss: 2.15651307371445e-05
step: 70, loss: 4.063825326738879e-05
step: 80, loss: 1.9058228645008057e-05
step: 90, loss: 0.003985458053648472
step: 100, loss: 1.554546906845644e-05
step: 110, loss: 3.618217306211591e-05
step: 120, loss: 2.5361210646224208e-05
step: 130, loss: 0.0016235650982707739
step: 140, loss: 2.439977106405422e-05
step: 150, loss: 2.731285167101305e-05
step: 160, loss: 1.9099266864941455e-05
step: 170, loss: 0.0784626230597496
step: 180, loss: 2.661253347469028e-05
step: 190, loss: 2.4373892301809974e-05
step: 200, loss: 0.0026630479842424393
step: 210, loss: 0.0018066921038553119
step: 220, loss: 0.00020453667093534023
step: 230, loss: 7.752649253234267e-05
step: 240, loss: 3.104516872554086e-05
step: 250, loss: 0.08274269849061966
step: 260, loss: 5.753750156145543e-05
step: 270, loss: 3.193158045178279e-05
step: 280, loss: 1.5225015886244364e-05
```

```
step: 290, loss: 2.4433426005998626e-05
step: 300, loss: 3.558456228347495e-05
step: 310, loss: 5.0434653530828655e-05
step: 320, loss: 3.009102147188969e-05
step: 330, loss: 6.343652785290033e-05
step: 340, loss: 2.211661012552213e-05
step: 350, loss: 0.07289052754640579
step: 360, loss: 1.78066056832904e-05
step: 370, loss: 1.7791726349969395e-05
step: 380, loss: 9.247809794032946e-05
step: 390, loss: 4.554781480692327e-05
step: 400, loss: 2.0864492398686707e-05
step: 410, loss: 2.6023893951787613e-05
step: 420, loss: 1.9721313947229646e-05
step: 430, loss: 0.0005716921295970678
step: 440, loss: 4.702278602053411e-05
step: 450, loss: 3.7604590033879504e-05
step: 460, loss: 1.7597960322746076e-05
step: 470, loss: 0.02386249601840973
step: 480, loss: 0.00221013720147311/
step: 490, loss: 3.2277690479531884e-05
step: 500, loss: 2.4087263227556832e-05
step: 510, loss: 2.330113056814298e-05
step: 520, loss: 4.2448376916581765e-05
step: 530, loss: 2.243684502900578e-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=0.9417337754618664, f1=0.940009447331129,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.3431464796885848e-05
step: 10, loss: 3.2910131267271936e-05
step: 20, loss: 1.4144785382086411e-05
step: 30, loss: 3.466183625278063e-05
step: 40, loss: 2.3505412173108198e-05
step: 50, loss: 4.638968312065117e-05
step: 60, loss: 1.8138185623683967e-05
step: 70, loss: 1.674118175287731e-05
step: 80, loss: 0.000290556694380939
step: 90, loss: 1.366795004287269e-05
step: 100, loss: 4.6134155127219856e-05
step: 110, loss: 1.6879052054719068e-05
step: 120, loss: 1.7407852283213288e-05
step: 130, loss: 5.336989852366969e-05
step: 140, loss: 0.0006611035787500441
```

```
step: 150, loss: 0.00012122141197323799
step: 160, loss: 0.0012045289622619748
step: 170, loss: 3.930835737264715e-05
step: 180, loss: 3.7049037928227335e-05
step: 190, loss: 0.0001682656438788399
step: 200, loss: 2.2496611563838087e-05
step: 210, loss: 0.00020128067990299314
step: 220, loss: 2.023913839366287e-05
step: 230, loss: 0.0028197960928082466
step: 240, loss: 3.225534237571992e-05
step: 250, loss: 0.0002740723721217364
step: 260, loss: 1.3917527212470304e-05
step: 270, loss: 0.003092344617471099
step: 280, loss: 5.880037497263402e-05
step: 290, loss: 2.846562711056322e-05
step: 300, loss: 2.8391799787641503e-05
step: 310, loss: 1.9717275790753774e-05
step: 320, loss: 1.8797518350766604e-05
step: 330, loss: 7.250000635394827e-05
step: 340, loss: 4.9940987082663924e-05
step: 350, loss: 9.465695620747283e-05
step: 360, loss: 0.002501817885786295
step: 370, loss: 6.172981375129893e-05
step: 380, loss: 0.0004366164794191718
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 390, loss: 1.8480832295608707e-05
step: 400, loss: 2.2284213628154248e-05
step: 410, loss: 2.474193570378702e-05
step: 420, loss: 1.561251156090293e-05
step: 430, loss: 1.764264925441239e-05
step: 440, loss: 1.4394357094715815e-05
step: 450, loss: 1.569445703353267e-05
step: 460, loss: 8.203316974686459e-05
step: 470, loss: 1.4364551134349313e-05
step: 480, loss: 2.1498066416825168e-05
step: 490, loss: 1.975450140889734e-05
step: 500, loss: 2.235480133094825e-05
step: 510, loss: 1.9166289348504506e-05
step: 520, loss: 2.438508454360999e-05
step: 530, loss: 2.397527714492753e-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=0.9416666666666667, f1=0.9441624365482235,
best_f1=0.9425815342214057
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 2.213845436926931e-05
step: 10, loss: 9.897423296933994e-05
step: 20, loss: 2.6366305974079296e-05
step: 30, loss: 2.851987665053457e-05
step: 40, loss: 2.1479178030858748e-05
step: 50, loss: 0.00020453172328416258
step: 60, loss: 1.6167492503882386e-05
step: 70, loss: 0.00042442657286301255
step: 80, loss: 2.162082091672346e-05
step: 90, loss: 1.3690287232748233e-05
step: 100, loss: 0.0001688292104518041
step: 110, loss: 0.02898041345179081
step: 120, loss: 0.0021108195651322603
step: 130, loss: 1.8104441551258788e-05
step: 140, loss: 1.436824823031202e-05
step: 150, loss: 0.009362388402223587
step: 160, loss: 1.826104926294647e-05
step: 170, loss: 2.6272664399584755e-05
step: 180, loss: 2.449626845191233e-05
step: 190, loss: 0.0005497168167494237
step: 200, loss: 4.387588705867529e-05
step: 210, loss: 2.959844277938828e-05
step: 220, loss: 1.1958081813645549e-05
step: 230, loss: 5.2807135944021866e-05
step: 240, loss: 2.0451427189982496e-05
step: 250, loss: 3.460099105723202e-05
step: 260, loss: 2.433972076687496e-05
step: 270, loss: 0.0026273068506270647
step: 280, loss: 3.110766556346789e-05
step: 290, loss: 0.0020413717720657587
step: 300, loss: 8.884411363396794e-05
step: 310, loss: 0.0002023845590883866
step: 320, loss: 5.395847620093264e-05
step: 330, loss: 2.7532205422176048e-05
step: 340, loss: 4.897038888884708e-05
step: 350, loss: 2.43281974690035e-05
step: 360, loss: 4.678622281062417e-05
step: 370, loss: 1.382812115480192e-05
step: 380, loss: 2.19111191652179696e-05
step: 390, loss: 1.6774709365563467e-05
step: 400, loss: 1.353013067273423e-05
step: 410, loss: 0.00012413815420586616
step: 420, loss: 0.000182799733011052
step: 430, loss: 2.0455161575227976e-05
step: 440, loss: 3.2445372198708355e-05
step: 450, loss: 0.00030415173387154937
step: 460, loss: 1.5537925719399936e-05
step: 470, loss: 3.648385245469399e-05
```

```
step: 480, loss: 0.0005297029856592417
step: 490, loss: 0.0003271025198046118
step: 500, loss: 7.749401265755296e-05
step: 510, loss: 0.00024157376901712269
step: 520, loss: 2.0857361960224807e-05
step: 530, loss: 2.3576481908094138e-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=0.9414519906323185, f1=0.9418874941887495,
best_f1=0.9425815342214057
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:13, 414.23it/s]
load_f1 = 0.9406350667280258
real_f1 = 0.9405803777061263
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 419.07it/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 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: 0.8231506943702698
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06559819728136063
step: 20, loss: 0.3752373456954956
step: 30, loss: 0.3760290741920471
step: 40, loss: 0.5064557790756226
step: 50, loss: 0.29842138290405273
step: 60, loss: 0.35366374254226685
step: 70, loss: 0.20332659780979156
step: 80, loss: 0.2947063148021698
step: 90, loss: 0.35229408740997314
step: 100, loss: 0.11317723989486694
step: 110, loss: 0.28380468487739563
step: 120, loss: 0.1995156854391098
step: 130, loss: 0.1812945008277893
step: 140, loss: 0.1924237608909607
step: 150, loss: 0.3370477557182312
step: 160, loss: 0.154353529214859
step: 170, loss: 0.12174669653177261
step: 180, loss: 0.13945485651493073
step: 190, loss: 0.22487126290798187
step: 200, loss: 0.13056784868240356
step: 210, loss: 0.42184969782829285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.631578947368421, f1=0.7116564417177913,
best_f1=0.7116564417177913
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05617798492312431
step: 10, loss: 0.11155981570482254
step: 20, loss: 0.15176938474178314
step: 30, loss: 0.1897696703672409
step: 40, loss: 0.07936914265155792
step: 50, loss: 0.20374152064323425
step: 60, loss: 0.045221228152513504
step: 70, loss: 0.13792872428894043
step: 80, loss: 0.13967633247375488
step: 90, loss: 0.12668144702911377
step: 100, loss: 0.15701313316822052
step: 110, loss: 0.07960618287324905
step: 120, loss: 0.17824330925941467
step: 130, loss: 0.1404159963130951
step: 140, loss: 0.13092227280139923
step: 150, loss: 0.10902362316846848
step: 160, loss: 0.1435372233390808
step: 170, loss: 0.1627800017595291
step: 180, loss: 0.18802350759506226
step: 190, loss: 0.10697272419929504
step: 200, loss: 0.1308290958404541
step: 210, loss: 0.1190212219953537
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6838966202783301, f1=0.7287128712871287,
best_f1=0.7287128712871287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11357170343399048
step: 10, loss: 0.17041617631912231
step: 20, loss: 0.2615782618522644
step: 30, loss: 0.057789456099271774
```

```
step: 40, loss: 0.16582143306732178
step: 50, loss: 0.12440931797027588
step: 60, loss: 0.07664172351360321
step: 70, loss: 0.07783285528421402
step: 80, loss: 0.02215002104640007
step: 90, loss: 0.08639586716890335
step: 100, loss: 0.06820444017648697
step: 110, loss: 0.05606765300035477
step: 120, loss: 0.12524722516536713
step: 130, loss: 0.06808746606111526
step: 140, loss: 0.2968449890613556
step: 150, loss: 0.12166734784841537
step: 160, loss: 0.07040669023990631
step: 170, loss: 0.13567492365837097
step: 180, loss: 0.0397341251373291
step: 190, loss: 0.39546096324920654
step: 200, loss: 0.12839782238006592
step: 210, loss: 0.09950247406959534
/usr/local/lib/python3.7/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=0.6702508960573477, f1=0.6842105263157895,
best_f1=0.7287128712871287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06969527900218964
step: 10, loss: 0.014001124538481236
step: 20, loss: 0.0705922394990921
step: 30, loss: 0.09696891903877258
step: 40, loss: 0.04065960273146629
step: 50, loss: 0.07005751132965088
step: 60, loss: 0.06787998974323273
step: 70, loss: 0.05561516433954239
step: 80, loss: 0.09966567903757095
step: 90, loss: 0.009837763383984566
step: 100, loss: 0.026902880519628525
step: 110, loss: 0.1358852982521057
step: 120, loss: 0.03301313519477844
step: 130, loss: 0.08732599020004272
step: 140, loss: 0.26890283823013306
step: 150, loss: 0.0780702754855156
step: 160, loss: 0.034332096576690674
step: 170, loss: 0.009702724404633045
step: 180, loss: 0.21540360152721405
step: 190, loss: 0.03938467428088188
step: 200, loss: 0.05251310393214226
step: 210, loss: 0.030499909073114395
```

```
/usr/local/lib/python3.7/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=0.6735112936344969, f1=0.6965376782077393,
best_f1=0.7287128712871287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05901089683175087
step: 10, loss: 0.06213642284274101
step: 20, loss: 0.018690941855311394
step: 30, loss: 0.10815369337797165
step: 40, loss: 0.0625695139169693
step: 50, loss: 0.05029958486557007
step: 60, loss: 0.060459088534116745
step: 70, loss: 0.04440867900848389
step: 80, loss: 0.004671365022659302
step: 90, loss: 0.05349086597561836
step: 100, loss: 0.013176998123526573
step: 110, loss: 0.009063346311450005
step: 120, loss: 0.005602733697742224
step: 130, loss: 0.01366925984621048
step: 140, loss: 0.13271211087703705
step: 150, loss: 0.05513341352343559
step: 160, loss: 0.008316985331475735
step: 170, loss: 0.0501287467777729
step: 180, loss: 0.038290269672870636
step: 190, loss: 0.0911862850189209
step: 200, loss: 0.11523009091615677
step: 210, loss: 0.010070471093058586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6861598440545809, f1=0.7013487475915221,
best_f1=0.7013487475915221
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13313882052898407
step: 10, loss: 0.020773397758603096
step: 20, loss: 0.012415856122970581
step: 30, loss: 0.0981668010354042
step: 40, loss: 0.003185346256941557
step: 50, loss: 0.03475257009267807
```

```
step: 60, loss: 0.1493348479270935
step: 70, loss: 0.008177334442734718
step: 80, loss: 0.10945575684309006
step: 90, loss: 0.08860877901315689
step: 100, loss: 0.05844623222947121
step: 110, loss: 0.012213022448122501
step: 120, loss: 0.05644405633211136
step: 130, loss: 0.13232798874378204
step: 140, loss: 0.01337787602096796
step: 150, loss: 0.11988662928342819
step: 160, loss: 0.10092542320489883
step: 170, loss: 0.046504173427820206
step: 180, loss: 0.01310514286160469
step: 190, loss: 0.12321426719427109
step: 200, loss: 0.006146220024675131
step: 210, loss: 0.0345650240778923
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6907020872865275, f1=0.7306967984934086,
best_f1=0.7306967984934086
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008822598494589329
step: 10, loss: 0.009479494765400887
step: 20, loss: 0.01190702710300684
step: 30, loss: 0.013908020220696926
step: 40, loss: 0.05661482363939285
step: 50, loss: 0.0847216546535492
step: 60, loss: 0.19590629637241364
step: 70, loss: 0.0026984564028680325
step: 80, loss: 0.015283611603081226
step: 90, loss: 0.009507202543318272
step: 100, loss: 0.0017880453960970044
step: 110, loss: 0.09764541685581207
step: 120, loss: 0.051718100905418396
step: 130, loss: 0.004918599501252174
step: 140, loss: 0.0007188786403276026
step: 150, loss: 0.036771006882190704
step: 160, loss: 0.005066532175987959
step: 170, loss: 0.03555944561958313
step: 180, loss: 0.000561029592063278
step: 190, loss: 0.0018150251125916839
```

```
step: 200, loss: 0.01978370174765587
step: 210, loss: 0.01703135296702385
/usr/local/lib/python3.7/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=0.6666666666666666, f1=0.6968503937007874,
best_f1=0.7306967984934086
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006753770634531975
step: 10, loss: 0.00082937267143279331
step: 20, loss: 0.021683165803551674
step: 30, loss: 0.012986100278794765
step: 40, loss: 0.024887053295969963
step: 50, loss: 0.010014946572482586
step: 60, loss: 0.031147297471761703
step: 70, loss: 0.0014069952303543687
step: 80, loss: 0.009094479493796825
step: 90, loss: 0.10891557484865189
step: 100, loss: 0.0053186602890491486
step: 110, loss: 0.0025062912609428167
step: 120, loss: 0.010588627308607101
step: 130, loss: 0.003094686195254326
step: 140, loss: 0.007928775623440742
step: 150, loss: 0.03678367659449577
step: 160, loss: 0.08178722858428955
step: 170, loss: 0.004613213241100311
step: 180, loss: 0.06431461125612259
step: 190, loss: 0.019555380567908287
step: 200, loss: 0.0035360674373805523
step: 210, loss: 0.19052883982658386
/usr/local/lib/python3.7/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=0.6828282828282828, f1=0.7056451612903225,
best_f1=0.7306967984934086
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00650422926992178
step: 10, loss: 0.0061654900200665
step: 20, loss: 0.01859637349843979
step: 30, loss: 0.04098613187670708
step: 40, loss: 0.036807581782341
step: 50, loss: 0.0005551596987061203
step: 60, loss: 0.007206633221358061
step: 70, loss: 0.0031011924147605896
```

```
step: 80, loss: 0.006910766940563917
step: 90, loss: 0.033874962478876114
step: 100, loss: 0.01579602062702179
step: 110, loss: 0.027510032057762146
step: 120, loss: 0.018198253586888313
step: 130, loss: 0.009355026297271252
step: 140, loss: 0.0216753501445055
step: 150, loss: 0.004915202036499977
step: 160, loss: 0.0018406036542728543
step: 170, loss: 0.14252705872058868
step: 180, loss: 0.006106338929384947
step: 190, loss: 0.0018071747617796063
step: 200, loss: 0.011639886535704136
step: 210, loss: 0.03767682611942291
/usr/local/lib/python3.7/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=0.6733466933867736, f1=0.698989898989899,
best_f1=0.7306967984934086
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006449359003454447
step: 10, loss: 0.010418413206934929
step: 20, loss: 0.05563022941350937
step: 30, loss: 0.027344951406121254
step: 40, loss: 0.027825966477394104
step: 50, loss: 0.00022544841340277344
step: 60, loss: 0.00019070944108534604
step: 70, loss: 0.1630651205778122
step: 80, loss: 0.0001881171192508191
step: 90, loss: 0.06007779389619827
step: 100, loss: 0.0006985776708461344
step: 110, loss: 0.0006237333291210234
step: 120, loss: 0.002722587436437607
step: 130, loss: 0.0015483840834349394
step: 140, loss: 0.0010047192918136716
step: 150, loss: 0.02890746295452118
step: 160, loss: 0.016079599037766457
step: 170, loss: 0.008626248687505722
step: 180, loss: 0.0035224289167672396
step: 190, loss: 0.13466057181358337
step: 200, loss: 0.02499035745859146
step: 210, loss: 0.0109579311683774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
```

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=0.6933333333333335, f1=0.7164179104477613,
best_f1=0.7164179104477613
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01113519724458456
step: 10, loss: 0.03395514190196991
step: 20, loss: 0.0004539993533398956
step: 30, loss: 0.04500419646501541
step: 40, loss: 0.0050616515800356865
step: 50, loss: 0.005387952551245689
step: 60, loss: 0.0006291373865678906
step: 70, loss: 0.000792770937550813
step: 80, loss: 0.11859537661075592
step: 90, loss: 0.0034703495912253857
step: 100, loss: 0.02219276875257492
step: 110, loss: 0.005910445936024189
step: 120, loss: 0.0011973055079579353
step: 130, loss: 0.026983071118593216
step: 140, loss: 0.005565640516579151
step: 150, loss: 0.041146691888570786
step: 160, loss: 0.003895187983289361
step: 170, loss: 0.06805694103240967
step: 180, loss: 0.002938389778137207
step: 190, loss: 0.0020224282052367926
step: 200, loss: 0.0002000626118388027
step: 210, loss: 0.0004523165989667177
/usr/local/lib/python3.7/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=0.6877470355731226, f1=0.7181467181467182,
best_f1=0.7164179104477613
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007018514443188906
step: 10, loss: 0.0002422466204734519
step: 20, loss: 0.004235150758177042
step: 30, loss: 0.0013936219038441777
step: 40, loss: 0.0019938619807362556
step: 50, loss: 0.01728604920208454
step: 60, loss: 0.013612361624836922
step: 70, loss: 0.00042882998241111636
step: 80, loss: 0.0015992746921256185
step: 90, loss: 0.0002729335392359644

```
step: 100, loss: 0.00040749291656538844
step: 110, loss: 0.00026251867529936135
step: 120, loss: 0.001249759690836072
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.01463998667895794
step: 140, loss: 0.13558533787727356
step: 150, loss: 0.0034924487117677927
step: 160, loss: 0.011582652106881142
step: 170, loss: 0.009744573384523392
step: 180, loss: 0.003114905674010515
step: 190, loss: 0.07912047952413559
step: 200, loss: 0.06327924132347107
step: 210, loss: 0.0016721525462344289
/usr/local/lib/python3.7/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=0.6751054852320675, f1=0.7056367432150314,
best_f1=0.7164179104477613
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000202293759007472545
step: 10, loss: 0.0009611102868802845
step: 20, loss: 0.0003284559352323413
step: 30, loss: 0.036935579031705856
step: 40, loss: 0.00020374001178424805
step: 50, loss: 0.0032047939021140337
step: 60, loss: 0.0175621435046196
step: 70, loss: 0.007728991564363241
step: 80, loss: 0.011160323396325111
step: 90, loss: 0.023826835677027702
step: 100, loss: 0.08607261627912521
step: 110, loss: 0.0008234078995883465
step: 120, loss: 0.004908304195851088
step: 130, loss: 0.0002894853241741657
step: 140, loss: 0.0017018287908285856
step: 150, loss: 0.00028306301101110876
step: 160, loss: 0.011802999302744865
step: 170, loss: 0.006917879451066256
step: 180, loss: 0.0859173834323883
step: 190, loss: 0.0005127700278535485
step: 200, loss: 0.0002618883445393294
step: 210, loss: 0.022211410105228424
/usr/local/lib/python3.7/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=0.6639175257731958, f1=0.7016129032258065,
best_f1=0.7164179104477613
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013456019572913647
step: 10, loss: 0.004141901154071093
step: 20, loss: 0.00016525383398402482
step: 30, loss: 0.00020209218200761825
step: 40, loss: 0.001888900762423873
step: 50, loss: 0.004061177372932434
step: 60, loss: 0.0026599441189318895
step: 70, loss: 0.007191620767116547
step: 80, loss: 0.04573597386479378
step: 90, loss: 0.0012095103738829494
step: 100, loss: 0.00033021910348907113
step: 110, loss: 0.007025541737675667
step: 120, loss: 0.0016620017122477293
step: 130, loss: 0.0005351725267246366
step: 140, loss: 0.00015179047477431595
step: 150, loss: 0.0006560657639056444
step: 160, loss: 0.0003038305731024593
step: 170, loss: 0.0010357089340686798
step: 180, loss: 0.007366083562374115
step: 190, loss: 0.00027289436548016965
step: 200, loss: 0.00018433878722134978
step: 210, loss: 0.008616620674729347
/usr/local/lib/python3.7/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=0.6772908366533865, f1=0.7134502923976609,
best_f1=0.7164179104477613
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0025313622318208218
step: 10, loss: 0.002902149921283126
step: 20, loss: 0.0008335871971212327
step: 30, loss: 0.00021120462042745203
step: 40, loss: 0.0004099093785043806
step: 50, loss: 0.00016049212717916816
step: 60, loss: 0.010342974215745926
step: 70, loss: 0.00022612823522649705
step: 80, loss: 0.000387163192499429
step: 90, loss: 0.05764163285493851
step: 100, loss: 0.0009662715601734817
step: 110, loss: 9.478213178226724e-05
step: 120, loss: 0.0018708924762904644
step: 130, loss: 0.03517882525920868
step: 140, loss: 0.0010586134158074856
```

```
step: 150, loss: 0.0033575412817299366
step: 160, loss: 0.0014662151224911213
step: 170, loss: 0.02501644566655159
step: 180, loss: 0.049042776226997375
step: 190, loss: 0.0009613941656425595
step: 200, loss: 0.14193713665008545
step: 210, loss: 0.03735462576150894
/usr/local/lib/python3.7/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=0.6720977596741344, f1=0.7080000000000001,
best_f1=0.7164179104477613
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.")
2293it [00:03, 759.07it/s]
load_f1 = 0.6666666666666666
real_f1 = 0.6714801444043321
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
4398it [00:10, 431.36it/s]
```

3.1.5 Walmart-Amazon - Running the matcher

```
[9]: # 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/
```

```
Downloading: 100% 442/442 [00:00<00:00, 493kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 2.66MB/s]
Downloading: 100% 268M/268M [00:03<00:00, 73.1MB/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
```

297

```
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: 0.847403347492218
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1615907996892929
step: 20, loss: 0.15511220693588257
step: 30, loss: 0.5172410607337952
step: 40, loss: 0.26864689588546753
step: 50, loss: 0.30703049898147583
step: 60, loss: 0.35960400104522705
step: 70, loss: 0.18186575174331665
step: 80, loss: 0.5594544410705566
step: 90, loss: 0.25394585728645325
step: 100, loss: 0.22277392446994781
step: 110, loss: 0.23828288912773132
step: 120, loss: 0.41865506768226624
step: 130, loss: 0.3391610085964203
step: 140, loss: 0.30972638726234436
step: 150, loss: 0.279349684715271
step: 160, loss: 0.1899491250514984
step: 170, loss: 0.335102379322052

```

```
step: 180, loss: 0.2654508948326111
step: 190, loss: 0.08998515456914902
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6391752577319586, f1=0.6461538461538462,
best_f1=0.6461538461538462
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.34321510791778564
step: 10, loss: 0.03271546959877014
step: 20, loss: 0.06348147243261337
step: 30, loss: 0.17047321796417236
step: 40, loss: 0.4095083177089691
step: 50, loss: 0.2962329387664795
step: 60, loss: 0.1583881378173828
step: 70, loss: 0.20944233238697052
step: 80, loss: 0.11804180592298508
step: 90, loss: 0.08618967235088348
step: 100, loss: 0.31962618231773376
step: 110, loss: 0.06624056398868561
step: 120, loss: 0.154342919588089
step: 130, loss: 0.15719562768936157
step: 140, loss: 0.05314909666776657
step: 150, loss: 0.01483608316630125
step: 160, loss: 0.018950581550598145
step: 170, loss: 0.04485602304339409
step: 180, loss: 0.07979503273963928
step: 190, loss: 0.13047349452972412
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7227722772277227, f1=0.7333333333333334,
best_f1=0.7333333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12957686185836792
step: 10, loss: 0.3174102008342743
```

```
step: 20, loss: 0.02342161536216736
step: 30, loss: 0.0706227719783783
step: 40, loss: 0.01350755337625742
step: 50, loss: 0.022440286353230476
step: 60, loss: 0.020087258890271187
step: 70, loss: 0.08327655494213104
step: 80, loss: 0.31916430592536926
step: 90, loss: 0.16055843234062195
step: 100, loss: 0.10418830066919327
step: 110, loss: 0.23895037174224854
step: 120, loss: 0.05677996203303337
step: 130, loss: 0.01698467507958412
step: 140, loss: 0.01843527890741825
step: 150, loss: 0.046982113271951675
step: 160, loss: 0.21529576182365417
step: 170, loss: 0.041924718767404556
step: 180, loss: 0.042632002383470535
step: 190, loss: 0.03169289976358414
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7810026385224274, f1=0.7733333333333333,
best_f1=0.7733333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0784367248415947
step: 10, loss: 0.02395625039935112
step: 20, loss: 0.009342280216515064
step: 30, loss: 0.03037775494158268
step: 40, loss: 0.0030824688728898764
step: 50, loss: 0.05246439948678017
step: 60, loss: 0.12889154255390167
step: 70, loss: 0.015038357116281986
step: 80, loss: 0.059441227465867996
step: 90, loss: 0.01724332757294178
step: 100, loss: 0.010583542287349701
step: 110, loss: 0.07058306783437729
step: 120, loss: 0.05944700539112091
step: 130, loss: 0.2798320949077606
step: 140, loss: 0.014274890534579754
step: 150, loss: 0.031659066677093506
step: 160, loss: 0.04714086279273033
step: 170, loss: 0.011716769076883793
```

```
step: 180, loss: 0.07510849088430405
step: 190, loss: 0.02301243133842945
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7846153846153846, f1=0.7552083333333333,
best_f1=0.7552083333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07969743013381958
step: 10, loss: 0.014188393950462341
step: 20, loss: 0.06816229224205017
step: 30, loss: 0.021170958876609802
step: 40, loss: 0.021503832191228867
step: 50, loss: 0.1628078818321228
step: 60, loss: 0.050025902688503265
step: 70, loss: 0.006834730505943298
step: 80, loss: 0.0054111173376441
step: 90, loss: 0.0018151372205466032
step: 100, loss: 0.011460642330348492
step: 110, loss: 0.0034182488452643156
step: 120, loss: 0.0015153206186369061
step: 130, loss: 0.0020665074698626995
step: 140, loss: 0.0026750604156404734
step: 150, loss: 0.014303538016974926
step: 160, loss: 0.060899410396814346
step: 170, loss: 0.021602747961878777
step: 180, loss: 0.10339582711458206
step: 190, loss: 0.09632906317710876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7935656836461127, f1=0.7500000000000001,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027121342718601227
step: 10, loss: 0.0012574098072946072
```

```
step: 20, loss: 0.0083650853484869
step: 30, loss: 0.003909592051059008
step: 40, loss: 0.09463731199502945
step: 50, loss: 0.002682870952412486
step: 60, loss: 0.0058102672919631
step: 70, loss: 0.0037824020255357027
step: 80, loss: 0.16787204146385193
step: 90, loss: 0.06428726762533188
step: 100, loss: 0.045664362609386444
step: 110, loss: 0.005177141632884741
step: 120, loss: 0.11050356179475784
step: 130, loss: 0.002361858496442437
step: 140, loss: 0.0023996769450604916
step: 150, loss: 0.02916555479168892
step: 160, loss: 0.04544112831354141
step: 170, loss: 0.009058114141225815
step: 180, loss: 0.00465022586286068
step: 190, loss: 0.007538771256804466
/usr/local/lib/python3.7/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=0.7703703703703704, f1=0.7438423645320198,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011293210554867983
step: 10, loss: 0.010217537172138691
step: 20, loss: 0.1520235389471054
step: 30, loss: 0.014552375301718712
step: 40, loss: 0.002540526445955038
step: 50, loss: 0.07225058972835541
step: 60, loss: 0.06916067004203796
step: 70, loss: 0.005673938430845737
step: 80, loss: 0.003201402025297284
step: 90, loss: 0.018506236374378204
step: 100, loss: 0.00572288827970624
step: 110, loss: 0.002244124421849847
step: 120, loss: 0.0010946899419650435
step: 130, loss: 0.0011483861599117517
step: 140, loss: 0.005760117899626493
step: 150, loss: 0.0025745427701622248
step: 160, loss: 0.0037041599862277508
step: 170, loss: 0.009545031934976578
step: 180, loss: 0.11149683594703674
step: 190, loss: 0.0009476927225477993
/usr/local/lib/python3.7/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=0.7594936708860759, f1=0.7684478371501272,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013530602445825934
step: 10, loss: 0.00257198978215456
step: 20, loss: 0.0016469318652525544
step: 30, loss: 0.012287036515772343
step: 40, loss: 0.0026385518722236156
step: 50, loss: 0.0013486410025507212
step: 60, loss: 0.0013216823572292924
step: 70, loss: 0.0012303043622523546
step: 80, loss: 0.028753293678164482
step: 90, loss: 0.001771487994119525
step: 100, loss: 0.020170658826828003
step: 110, loss: 0.005342619493603706
step: 120, loss: 0.0004281859437469393
step: 130, loss: 0.005749562289565802
step: 140, loss: 0.0006370680057443678
step: 150, loss: 0.0006542964838445187
step: 160, loss: 0.00047319120494648814
step: 170, loss: 0.0008562011062167585
step: 180, loss: 0.018548298627138138
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 190, loss: 0.02590341493487358
/usr/local/lib/python3.7/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=0.7814207650273224, f1=0.7675675675675676,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000475471984827891
step: 10, loss: 0.0012824749574065208
step: 20, loss: 0.1032683327794075
step: 30, loss: 0.001645198673941195
step: 40, loss: 0.0070885862223804
step: 50, loss: 0.0016016490990296006
step: 60, loss: 0.04986179992556572
step: 70, loss: 0.002341472776606679
step: 80, loss: 0.010800746269524097
step: 90, loss: 0.010657218284904957
step: 100, loss: 0.0022198548540472984
step: 110, loss: 0.0037111996207386255
step: 120, loss: 0.012208592146635056
```

```
step: 130, loss: 0.0034711204934865236
step: 140, loss: 0.0007038495969027281
step: 150, loss: 0.007212987635284662
step: 160, loss: 0.0014429192524403334
step: 170, loss: 0.02073185332119465
step: 180, loss: 0.03198694437742233
step: 190, loss: 0.0042852722108364105
/usr/local/lib/python3.7/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=0.7747252747252747, f1=0.7507002801120448,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006885752663947642
step: 10, loss: 0.007999484427273273
step: 20, loss: 0.0013937036274001002
step: 30, loss: 0.0011357000330463052
step: 40, loss: 0.0003554338763933629
step: 50, loss: 0.0006197227630764246
step: 60, loss: 0.026651110500097275
step: 70, loss: 0.0018591736443340778
step: 80, loss: 0.0009355025831609964
step: 90, loss: 0.00160305411554873
step: 100, loss: 0.0003606936370488256
step: 110, loss: 0.031618934124708176
step: 120, loss: 0.005322339478880167
step: 130, loss: 0.0003020019503310323
step: 140, loss: 0.0003714062040671706
step: 150, loss: 0.00035173515789210796
step: 160, loss: 0.000829319644253701
step: 170, loss: 0.00041979251545853913
step: 180, loss: 0.00033269289997406304
step: 190, loss: 0.0005630181985907257
/usr/local/lib/python3.7/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=0.7862796833773087, f1=0.7716535433070866,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003642455267254263
step: 10, loss: 0.0007478052284568548
step: 20, loss: 0.0010645822621881962
step: 30, loss: 0.0027470870409160852
step: 40, loss: 0.0002582100569270551
```

```
step: 50, loss: 0.000337040051817894
step: 60, loss: 0.0008675004355609417
step: 70, loss: 0.0002522130380384624
step: 80, loss: 0.0001625317381694913
step: 90, loss: 0.002225010422989726
step: 100, loss: 0.002197163412347436
step: 110, loss: 0.00023087899899110198
step: 120, loss: 0.001265446306206286
step: 130, loss: 0.0015427835751324892
step: 140, loss: 0.0010928245028480887
step: 150, loss: 0.0009059259318746626
step: 160, loss: 0.0005342685035429895
step: 170, loss: 0.02423745021224022
step: 180, loss: 0.01062440313398838
step: 190, loss: 0.00035046847187913954
/usr/local/lib/python3.7/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=0.779291553133515, f1=0.7554347826086956,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023918552324175835
step: 10, loss: 0.0005852541653439403
step: 20, loss: 0.0011910123284906149
step: 30, loss: 0.0051099746488034725
step: 40, loss: 0.013425429351627827
step: 50, loss: 0.0036429534666240215
step: 60, loss: 0.0002214980049757287
step: 70, loss: 0.000550299184396863
step: 80, loss: 0.0004711384535767138
step: 90, loss: 0.11838512122631073
step: 100, loss: 0.0007868465036153793
step: 110, loss: 0.005434572231024504
step: 120, loss: 0.00024781899992376566
step: 130, loss: 0.00014036198263056576
step: 140, loss: 0.005353442393243313
step: 150, loss: 0.0005657229339703918
step: 160, loss: 0.049555663019418716
step: 170, loss: 0.023258376866579056
step: 180, loss: 0.005631447769701481
step: 190, loss: 0.00614667683839798
/usr/local/lib/python3.7/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=0.7627118644067796, f1=0.7570621468926554,
best_f1=0.7500000000000001
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013679104158654481
step: 10, loss: 0.0016707482282072306
step: 20, loss: 0.012827988713979721
step: 30, loss: 0.00024885626044124365
step: 40, loss: 0.0005850537563674152
step: 50, loss: 0.0008465218706987798
step: 60, loss: 0.00022449342941399664
step: 70, loss: 0.0006037374259904027
step: 80, loss: 0.00047852846910245717
step: 90, loss: 0.025268107652664185
step: 100, loss: 0.0002742784272413701
step: 110, loss: 0.00048087415052577853
step: 120, loss: 0.002730102278292179
step: 130, loss: 0.001028272439725697
step: 140, loss: 0.00026060492382384837
step: 150, loss: 0.0005417291540652514
step: 160, loss: 0.0004795854911208153
step: 170, loss: 0.00015153760614339262
step: 180, loss: 0.0664137527346611
step: 190, loss: 0.009119651280343533
/usr/local/lib/python3.7/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=0.761904761904762, f1=0.7454068241469817,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001720645697787404
step: 10, loss: 0.0014812520239502192
step: 20, loss: 0.0006809812039136887
step: 30, loss: 0.01271249633282423
step: 40, loss: 0.00014546093007083982
step: 50, loss: 0.0017211916856467724
step: 60, loss: 0.00014374589954968542
step: 70, loss: 0.0006583150825463235
step: 80, loss: 9.301812679041177e-05
step: 90, loss: 0.0006934224511496723
step: 100, loss: 0.0005688399542123079
step: 110, loss: 0.00021449968335218728
step: 120, loss: 0.00031838740687817335
step: 130, loss: 0.0002754631859716028
step: 140, loss: 0.00021596584701910615
step: 150, loss: 0.0003425291506573558
step: 160, loss: 0.0002263823407702148
```

```
step: 170, loss: 0.00024829263566061854
step: 180, loss: 0.001223166473209858
step: 190, loss: 0.00018595557776279747
/usr/local/lib/python3.7/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=0.774869109947644, f1=0.7461139896373058,
best_f1=0.7500000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033113022800534964
step: 10, loss: 0.0009844216983765364
step: 20, loss: 0.0009861659491434693
step: 30, loss: 0.0001578425581101328
step: 40, loss: 0.0004923765663988888
step: 50, loss: 0.03021906688809395
step: 60, loss: 0.00035232160007581115
step: 70, loss: 0.00020617470727302134
step: 80, loss: 0.00010883025970542803
step: 90, loss: 0.006812622305005789
step: 100, loss: 0.0007773301913402975
step: 110, loss: 0.00026678931317292154
step: 120, loss: 0.0017552203498780727
step: 130, loss: 0.0023220579605549574
step: 140, loss: 0.0003680622612591833
step: 150, loss: 0.0003981452900916338
step: 160, loss: 0.0005661037284880877
step: 170, loss: 0.00032043890678323805
step: 180, loss: 0.00024699122877791524
step: 190, loss: 0.016599826514720917
/usr/local/lib/python3.7/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=0.773841961852861, f1=0.7580645161290323,
best_f1=0.7500000000000001
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 [00:05, 385.54it/s]
load_f1 = 0.5942028985507246
real_f1 = 0.5843230403800476
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 [00:09, 442.58it/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 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: 0.851843535900116
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22627083957195282
step: 20, loss: 0.15417855978012085
step: 30, loss: 0.22878757119178772
step: 40, loss: 0.31910544633865356
step: 50, loss: 0.37830474972724915
step: 60, loss: 0.44554346799850464
step: 70, loss: 0.3148956298828125
step: 80, loss: 0.24698875844478607
step: 90, loss: 0.40499722957611084
step: 100, loss: 0.23829831182956696
step: 110, loss: 0.18966449797153473
step: 120, loss: 0.5407047867774963
step: 130, loss: 0.41106879711151123
step: 140, loss: 0.48486801981925964
step: 150, loss: 0.11016889661550522
step: 160, loss: 0.24893774092197418
step: 170, loss: 0.1818351000547409
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6421663442940038, f1=0.63671875, best_f1=0.63671875
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2045692354440689
step: 10, loss: 0.0441044420003891
step: 20, loss: 0.32308229804039
step: 30, loss: 0.25640812516212463
step: 40, loss: 0.1303676962852478
step: 50, loss: 0.18865175545215607
step: 60, loss: 0.0762481614947319
step: 70, loss: 0.15908709168434143
step: 80, loss: 0.10616397112607956
step: 90, loss: 0.11001483350992203
step: 100, loss: 0.1505989134311676
step: 110, loss: 0.13445556163787842
step: 120, loss: 0.05584179237484932

```
step: 130, loss: 0.0654078871011734
step: 140, loss: 0.10629042237997055
step: 150, loss: 0.09639491885900497
step: 160, loss: 0.11933420598506927
step: 170, loss: 0.0809214860200882
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7895981087470448, f1=0.7644444444444445,
best_f1=0.7644444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03299213573336601
step: 10, loss: 0.0310016218572855
step: 20, loss: 0.05136396363377571
step: 30, loss: 0.1861753910779953
step: 40, loss: 0.010732765309512615
step: 50, loss: 0.1852584034204483
step: 60, loss: 0.15650790929794312
step: 70, loss: 0.03685416281223297
step: 80, loss: 0.21875829994678497
step: 90, loss: 0.07991914451122284
step: 100, loss: 0.0272170826792716698
step: 110, loss: 0.028097691014409065
step: 120, loss: 0.012084110639989376
step: 130, loss: 0.1689210683107376
step: 140, loss: 0.02221066877245903
step: 150, loss: 0.0964491218328476
step: 160, loss: 0.03865738585591316
step: 170, loss: 0.16183938086032867
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8241206030150753, f1=0.7862407862407862,
best_f1=0.7862407862407862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.044350579380989075
```

```
step: 10, loss: 0.17201536893844604
step: 20, loss: 0.017689626663923264
step: 30, loss: 0.022335492074489594
step: 40, loss: 0.024234792217612267
step: 50, loss: 0.02502264827489853
step: 60, loss: 0.04052316024899483
step: 70, loss: 0.10116016864776611
step: 80, loss: 0.03965560719370842
step: 90, loss: 0.025544440373778343
step: 100, loss: 0.009009145200252533
step: 110, loss: 0.01699320413172245
step: 120, loss: 0.12957628071308136
step: 130, loss: 0.024842694401741028
step: 140, loss: 0.01858401484787464
step: 150, loss: 0.0039040485862642527
step: 160, loss: 0.03217948600649834
step: 170, loss: 0.0085292253643274
/usr/local/lib/python3.7/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=0.788659793814433, f1=0.8, best_f1=0.7862407862407862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017895514145493507
step: 10, loss: 0.015634002164006233
step: 20, loss: 0.011502050794661045
step: 30, loss: 0.08317040652036667
step: 40, loss: 0.14970111846923828
step: 50, loss: 0.0221692007035017
step: 60, loss: 0.0020976385567337275
step: 70, loss: 0.0262132678180933
step: 80, loss: 0.002911589341238141
step: 90, loss: 0.004057253245264292
step: 100, loss: 0.026430776342749596
step: 110, loss: 0.05175287276506424
step: 120, loss: 0.03789616376161575
step: 130, loss: 0.003633642802014947
step: 140, loss: 0.01677284762263298
step: 150, loss: 0.016978131607174873
step: 160, loss: 0.0023228293284773827
step: 170, loss: 0.004537865985184908
/usr/local/lib/python3.7/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=0.8008849557522124, f1=0.8059701492537314,
best_f1=0.7862407862407862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018354199826717377
step: 10, loss: 0.05338974669575691
step: 20, loss: 0.0033612512052059174
step: 30, loss: 0.06299997866153717
step: 40, loss: 0.006178438663482666
step: 50, loss: 0.010569511912763119
step: 60, loss: 0.06645742803812027
step: 70, loss: 0.0066696517169475555
step: 80, loss: 0.029009805992245674
step: 90, loss: 0.052838411182165146
step: 100, loss: 0.1062052994966507
step: 110, loss: 0.015579495579004288
step: 120, loss: 0.023459507152438164
step: 130, loss: 0.2006070762872696
step: 140, loss: 0.029293682426214218
step: 150, loss: 0.14497216045856476
step: 160, loss: 0.016422264277935028
step: 170, loss: 0.002931553404778242
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8254716981132074, f1=0.8216704288939052,
best_f1=0.8216704288939052
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007148898206651211
step: 10, loss: 0.004392060451209545
step: 20, loss: 0.05073005333542824
step: 30, loss: 0.0061554862186312675
step: 40, loss: 0.006692584604024887
step: 50, loss: 0.0010283506708219647
step: 60, loss: 0.09529274702072144
step: 70, loss: 0.0013827235670760274
step: 80, loss: 0.005260032135993242
step: 90, loss: 0.006930580362677574
step: 100, loss: 0.01136021688580513
step: 110, loss: 0.00040492526022717357
step: 120, loss: 0.029626326635479927
step: 130, loss: 0.007356039248406887
step: 140, loss: 0.07232364267110825
step: 150, loss: 0.05689071863889694
```

```
step: 160, loss: 0.0027627318631857634
step: 170, loss: 0.1459994614124298
/usr/local/lib/python3.7/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=0.803970223325062, f1=0.8045977011494253,
best_f1=0.8216704288939052
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01755974255502224
step: 10, loss: 0.009334148839116096
step: 20, loss: 0.004079263191670179
step: 30, loss: 0.02961164154112339
step: 40, loss: 0.001636614790186286
step: 50, loss: 0.0027414108626544476
step: 60, loss: 0.023585503920912743
step: 70, loss: 0.019164983183145523
step: 80, loss: 0.0022036307491362095
step: 90, loss: 0.014841298572719097
step: 100, loss: 0.007600439712405205
step: 110, loss: 0.07077674567699432
step: 120, loss: 0.0023885348346084356
step: 130, loss: 0.0006286471034400165
step: 140, loss: 0.00977829098701477
step: 150, loss: 0.052445389330387115
step: 160, loss: 0.021686699241399765
step: 170, loss: 0.002475558314472437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8291457286432161, f1=0.807785888077859,
best_f1=0.807785888077859
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005724302027374506
step: 10, loss: 0.04053455591201782
step: 20, loss: 0.011602955870330334
step: 30, loss: 0.047837771475315094
step: 40, loss: 0.013013313524425030
step: 50, loss: 0.09486815333366394
step: 60, loss: 0.002438528463244438
step: 70, loss: 0.009240606799721718
```

```
step: 80, loss: 0.03421814739704132
step: 90, loss: 0.1296442151069641
step: 100, loss: 0.06394161283969879
step: 110, loss: 0.002249169861897826
step: 120, loss: 0.008839774876832962
step: 130, loss: 0.010484401136636734
step: 140, loss: 0.1137782633304596
step: 150, loss: 0.009159069508314133
step: 160, loss: 0.046035025268793106
step: 170, loss: 0.005829433910548687
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8307692307692308, f1=0.8138957816377173,
best_f1=0.8138957816377173
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000924094463698566
step: 10, loss: 0.033399391919374466
step: 20, loss: 0.0011052257614210248
step: 30, loss: 0.0016372386598959565
step: 40, loss: 0.0015672363806515932
step: 50, loss: 0.0021287440322339535
step: 60, loss: 0.003088773926720023
step: 70, loss: 0.004732165951281786
step: 80, loss: 0.016792627051472664
step: 90, loss: 0.030355338007211685
step: 100, loss: 0.00450073042884469
step: 110, loss: 0.004635619930922985
step: 120, loss: 0.013350413180887699
step: 130, loss: 0.003461349057033658
step: 140, loss: 0.024096496403217316
step: 150, loss: 0.00028617813950404525
step: 160, loss: 0.0006165321101434529
step: 170, loss: 0.005779193248599768
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.85, f1=0.8221153846153845, best_f1=0.8221153846153845
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002438038500258699
step: 10, loss: 0.008630117401480675
step: 20, loss: 0.000761273258831352
step: 30, loss: 0.0006406921311281621
step: 40, loss: 0.013318156823515892
step: 50, loss: 0.01162816770374775
step: 60, loss: 0.006744707468897104
step: 70, loss: 0.002439378062263131
step: 80, loss: 0.0003565519582480192
step: 90, loss: 0.0044363122433424
step: 100, loss: 0.0003498440491966903
step: 110, loss: 0.001263663056306541
step: 120, loss: 0.0011407813290134072
step: 130, loss: 0.0004659423138946295
step: 140, loss: 0.021508747711777687
step: 150, loss: 0.002222393872216344
step: 160, loss: 0.008702694438397884
step: 170, loss: 0.11702173948287964
/usr/local/lib/python3.7/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=0.8317307692307692, f1=0.8264840182648402,
best_f1=0.8221153846153845
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006143689621239901
step: 10, loss: 0.0021751276217401028
step: 20, loss: 0.0120710963383317
step: 30, loss: 0.0005386778502725065
step: 40, loss: 0.00040954910218715567
step: 50, loss: 0.00021802762057632208
step: 60, loss: 0.00030098066781647503
step: 70, loss: 0.012893645092844963
step: 80, loss: 0.046568792313337326
step: 90, loss: 0.003111968282610178
step: 100, loss: 0.001163232373073697
step: 110, loss: 0.00016523903468623757
step: 120, loss: 0.0011823370587080717
step: 130, loss: 0.0003517670265864581
step: 140, loss: 0.0003843529266305268
step: 150, loss: 0.005386655684560537
step: 160, loss: 0.004121961537748575
step: 170, loss: 0.002719002077355981
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.")
epoch 12: dev_f1=0.8428571428571429, f1=0.8195991091314031,
best_f1=0.8221153846153845
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00025200058007612824
step: 10, loss: 0.08387162536382675
step: 20, loss: 0.00020094426872674376
step: 30, loss: 0.00016661525311646983
step: 40, loss: 0.0003077270230278373
step: 50, loss: 0.0011300885817036033
step: 60, loss: 0.00040143722435459495
step: 70, loss: 0.00034393888199701905
step: 80, loss: 0.0005557016702368855
step: 90, loss: 0.0019582121167331934
step: 100, loss: 0.00010642351116985083
step: 110, loss: 0.0002991500950884074
step: 120, loss: 0.0011968349572271109
step: 130, loss: 0.013522197492420673
step: 140, loss: 9.357866656500846e-05
step: 150, loss: 0.00017352003487758338
step: 160, loss: 0.00015136129513848573
step: 170, loss: 0.0004931093426421285
/usr/local/lib/python3.7/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=0.8447837150127225, f1=0.8195121951219512,
best_f1=0.8221153846153845
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002121024066582322
step: 10, loss: 0.003421484725549817
step: 20, loss: 0.01246603112667799
step: 30, loss: 0.0037124420050531626
step: 40, loss: 0.0008390223374590278
step: 50, loss: 0.0022670079488307238
step: 60, loss: 0.000415043585235253
step: 70, loss: 0.0007599074160680175
step: 80, loss: 0.008161542005836964
step: 90, loss: 0.00032092409674078226
step: 100, loss: 0.00029098481172695756
step: 110, loss: 0.0002948856563307345
step: 120, loss: 0.0446137972176075
step: 130, loss: 0.0002055098448181525
```

```
step: 140, loss: 0.0008243367774412036
step: 150, loss: 0.00024876362294889987
step: 160, loss: 0.02883654274046421
step: 170, loss: 0.00020153317018412054
/usr/local/lib/python3.7/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=0.8345323741007193, f1=0.8089887640449437,
best_f1=0.8221153846153845
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010693237418308854
step: 10, loss: 0.014364882372319698
step: 20, loss: 0.010214641690254211
step: 30, loss: 0.01100306399166584
step: 40, loss: 0.0002279917971463874
step: 50, loss: 0.0010604048147797585
step: 60, loss: 8.950640767579898e-05
step: 70, loss: 0.00036503831506706774
step: 80, loss: 0.0003448856878094375
step: 90, loss: 0.00013782794121652842
step: 100, loss: 0.0004529180296231061
step: 110, loss: 0.0001950725563801825
step: 120, loss: 0.0029402084182947874
step: 130, loss: 0.13490888476371765
step: 140, loss: 0.0013950483407825232
step: 150, loss: 0.029073458164930344
step: 160, loss: 0.00010047804971691221
step: 170, loss: 0.0011094786459580064
/usr/local/lib/python3.7/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=0.832535885167464, f1=0.8108108108108107,
best_f1=0.8221153846153845
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 [00:04, 463.83it/s]
load_f1 = 0.45911047345767575
real_f1 = 0.4628330995792427
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 420.88it/s]
```

3.3 DK DIRTY

3.3.1 DBLP-ACM - Running the matcher

```python
# 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 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: 0.7982252836227417
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.46126115322113037
step: 20, loss: 0.5793554186820984
step: 30, loss: 0.3891143798828125
step: 40, loss: 0.147667795419693
step: 50, loss: 0.129313662648201
step: 60, loss: 0.1234314814209938
step: 70, loss: 0.13944485783576965
step: 80, loss: 0.14700129628181458
step: 90, loss: 0.032267648726701736
step: 100, loss: 0.0733431875705719
step: 110, loss: 0.06259417533874512
step: 120, loss: 0.025171151384711266
step: 130, loss: 0.009950090199708939
step: 140, loss: 0.008292080834507942
step: 150, loss: 0.16763295233249664
step: 160, loss: 0.21932335197925568
step: 170, loss: 0.05130065232515335
step: 180, loss: 0.026951339095830917
step: 190, loss: 0.008552365005016327
step: 200, loss: 0.007559474557638168
step: 210, loss: 0.008734012022614479
step: 220, loss: 0.006240388378500938
step: 230, loss: 0.008046884089708328
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9694224235560589, f1=0.9622857142857143,
best_f1=0.9622857142857143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03796403110027313
step: 10, loss: 0.004777185153216124
step: 20, loss: 0.0024204535875469446
step: 30, loss: 0.00282215210609138
step: 40, loss: 0.01369047537446022
step: 50, loss: 0.017595632001757622
step: 60, loss: 0.07775889337062836

```
step: 70, loss: 0.14883625507354736
step: 80, loss: 0.005850037559866905
step: 90, loss: 0.0042567471973598
step: 100, loss: 0.1762108951807022
step: 110, loss: 0.15021684765815735
step: 120, loss: 0.0028711655177176
step: 130, loss: 0.0025516932364553213
step: 140, loss: 0.316579669713974
step: 150, loss: 0.030808674171566963
step: 160, loss: 0.06864020973443985
step: 170, loss: 0.021999729797244072
step: 180, loss: 0.011657536029815674
step: 190, loss: 0.17161619663238525
step: 200, loss: 0.09108994156122208
step: 210, loss: 0.08170433342456818
step: 220, loss: 0.002446193713694811
step: 230, loss: 0.005960991606116295
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9753914988814317, f1=0.9698324022346367,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04466325417160988
step: 10, loss: 0.006130996160209179
step: 20, loss: 0.005930757150053978
step: 30, loss: 0.04159817844629288
step: 40, loss: 0.02522071823477745
step: 50, loss: 0.016325989738106728
step: 60, loss: 0.0014155867975205183
step: 70, loss: 0.001371042220853269
step: 80, loss: 0.016728442162275314
step: 90, loss: 0.005332452245056629
step: 100, loss: 0.0012868251651525497
step: 110, loss: 0.008922464214265347
step: 120, loss: 0.014607506804168224
step: 130, loss: 0.006624886766076088
step: 140, loss: 0.0023954594507813454
step: 150, loss: 0.001534952549263835
step: 160, loss: 0.0025306222960352898
step: 170, loss: 0.009081744588911533
step: 180, loss: 0.004351045936346054
```

```
step: 190, loss: 0.01026318408548832
step: 200, loss: 0.005060279741883278
step: 210, loss: 0.010682199150323868
step: 220, loss: 0.009411705657839775
step: 230, loss: 0.1244627833366394
/usr/local/lib/python3.7/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=0.9743589743589743, f1=0.9733333333333333,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012831992469727993
step: 10, loss: 0.0016226621810346842
step: 20, loss: 0.008541599847376347
step: 30, loss: 0.003481904976069927
step: 40, loss: 0.0166498813778162
step: 50, loss: 0.00806766003370285
step: 60, loss: 0.0006220935028977692
step: 70, loss: 0.009341568686068058
step: 80, loss: 0.027727702632546425
step: 90, loss: 0.0019792960956692696
step: 100, loss: 0.00149274873547256
step: 110, loss: 0.015865005552768707
step: 120, loss: 0.010226459242403507
step: 130, loss: 0.0013205313589423895
step: 140, loss: 0.0005202881875447929
step: 150, loss: 0.0011591697111725807
step: 160, loss: 0.0021602369379252195
step: 170, loss: 0.05761796981096268
step: 180, loss: 0.041802700608968735
step: 190, loss: 0.0012787638697773218
step: 200, loss: 0.0011753594735637307
step: 210, loss: 0.0004474686284083873
step: 220, loss: 0.000736779416911304
step: 230, loss: 0.0007644834113307297
/usr/local/lib/python3.7/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=0.974472807991121, f1=0.9766407119021134,
best_f1=0.9698324022346367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008376339683309197
step: 10, loss: 0.0009385606390424073
step: 20, loss: 0.0009364433353766799
```

```
step: 30, loss: 0.00028637467767111957
step: 40, loss: 0.0007206816226243973
step: 50, loss: 0.0010990714654326439
step: 60, loss: 0.0008734612492844462
step: 70, loss: 0.00019431674445513636
step: 80, loss: 0.0005739984335377812
step: 90, loss: 0.0008622129098512232
step: 100, loss: 0.002948096487671137
step: 110, loss: 0.0005572201334871352
step: 120, loss: 0.0645802766084671
step: 130, loss: 0.028156409040093422
step: 140, loss: 0.00367254507727921
step: 150, loss: 0.008176383562386036
step: 160, loss: 0.14248771965503693
step: 170, loss: 0.05901600420475006
step: 180, loss: 0.009043904952704906
step: 190, loss: 0.0010902624344453216
step: 200, loss: 0.004380512051284313
step: 210, loss: 0.0027453771326690912
step: 220, loss: 0.0013058355543762445
step: 230, loss: 0.0009776903316378593
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9764837625979844, f1=0.9764309764309763,
best_f1=0.9764309764309763
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017493043560534716
step: 10, loss: 0.0003408904012758285
step: 20, loss: 0.0014533490175381303
step: 30, loss: 0.0018541570752859116
step: 40, loss: 0.0024217364843934774
step: 50, loss: 0.019083037972450256
step: 60, loss: 0.003461959073320031
step: 70, loss: 0.001374159473925829
step: 80, loss: 0.00044131974573247135
step: 90, loss: 0.00036986664053983986
step: 100, loss: 0.00046498983283527195
step: 110, loss: 0.018133876845240593
step: 120, loss: 0.00026340471231378615
step: 130, loss: 0.0006186716491356492
step: 140, loss: 0.00030580523889511824
```

```
step: 150, loss: 0.00023022993991617113
step: 160, loss: 0.0114821782335639
step: 170, loss: 0.0008118702680803835
step: 180, loss: 0.002158304676413536
step: 190, loss: 0.0073785195127129555
step: 200, loss: 0.009767038747668266
step: 210, loss: 0.0023062294349074364
step: 220, loss: 0.0005800877115689218
step: 230, loss: 0.0006562613416463137
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.977728285077951, f1=0.9723145071982282,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10907960683107376
step: 10, loss: 0.0003604549274314195
step: 20, loss: 0.0003629476123023778
step: 30, loss: 0.0013742103474214673
step: 40, loss: 0.009251924231648445
step: 50, loss: 0.0010462704813107848
step: 60, loss: 0.016390802338719368
step: 70, loss: 0.00020261104509700009
step: 80, loss: 0.00013400659372564405
step: 90, loss: 0.0010095578618347645
step: 100, loss: 0.00036965246545150876
step: 110, loss: 0.02194429561495781
step: 120, loss: 0.0014425779227167368
step: 130, loss: 0.0046775564551353455
step: 140, loss: 0.00018104554328601807
step: 150, loss: 0.0006356082740239799
step: 160, loss: 0.0005564826424233615
step: 170, loss: 0.0001334420667262748
step: 180, loss: 0.00022691635240335017
step: 190, loss: 0.0006467702332884073
step: 200, loss: 0.001122856861911714
step: 210, loss: 9.61787227424793e-05
step: 220, loss: 0.0015431210631504655
step: 230, loss: 0.04215914011001587
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.978912319644839, f1=0.9711751662971175,
best_f1=0.9711751662971175
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020275339484214783
step: 10, loss: 0.0065385401248931885
step: 20, loss: 0.002662436803802848
step: 30, loss: 0.00025157156051136553
step: 40, loss: 0.0005383603274822235
step: 50, loss: 0.003127357689663768
step: 60, loss: 0.0007943498203530908
step: 70, loss: 0.003601479111239314
step: 80, loss: 0.0014854173641651869
step: 90, loss: 0.0002067173772957176
step: 100, loss: 0.00040089740650728345
step: 110, loss: 0.0030890442430973053
step: 120, loss: 0.0001722770684864372
step: 130, loss: 0.0002723456418607384
step: 140, loss: 0.00046106221270747483
step: 150, loss: 0.00031333649531006813
step: 160, loss: 0.0005845740670338273
step: 170, loss: 0.0009849353227764368
step: 180, loss: 0.001050548511557281
step: 190, loss: 0.0007847534725442529
step: 200, loss: 0.013378328643739223
step: 210, loss: 0.00018913493840955198
step: 220, loss: 0.0007370912935584784
step: 230, loss: 0.042916182428598404
/usr/local/lib/python3.7/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=0.9763779527559054, f1=0.9753363228699552,
best_f1=0.9711751662971175
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.738614608068019e-05
step: 10, loss: 0.00030333123868331313
step: 20, loss: 0.0008474886417388916
step: 30, loss: 0.0004670759371947497
step: 40, loss: 6.0071419284213334e-05
step: 50, loss: 0.017652807757258415
step: 60, loss: 0.00020599516574293375
```

```
step: 70, loss: 0.0003271323803346604
step: 80, loss: 0.18790599703788757
step: 90, loss: 0.017381859943270683
step: 100, loss: 0.008936627767980099
step: 110, loss: 0.000372575712390244
step: 120, loss: 0.0678337886929512
step: 130, loss: 0.0005522509454749525
step: 140, loss: 0.09172425419092178
step: 150, loss: 6.487582140835002e-05
step: 160, loss: 0.0002438213850837201
step: 170, loss: 0.00010709337220760062
step: 180, loss: 0.0003452735545579344
step: 190, loss: 0.00012230934225954115
step: 200, loss: 0.00016481200873386115
step: 210, loss: 0.00010054435551865026
step: 220, loss: 0.05962619557976723
step: 230, loss: 0.037315405905246735
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9821428571428571, f1=0.9723145071982282,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011531300697242841
step: 10, loss: 0.00016422811313532293
step: 20, loss: 0.00033833840279839933
step: 30, loss: 0.00013570919691119343
step: 40, loss: 5.166902337805368e-05
step: 50, loss: 0.0031154609750956297
step: 60, loss: 0.08688604086637497
step: 70, loss: 0.0001056498585967347
step: 80, loss: 0.00020015069458168
step: 90, loss: 8.055302168941125e-05
step: 100, loss: 0.0004628051246982068
step: 110, loss: 0.0004402881895657629
step: 120, loss: 0.020462000742554665
step: 130, loss: 0.026882482692599297
step: 140, loss: 0.0038526570424437523
step: 150, loss: 0.0003751228505279869
step: 160, loss: 0.0003448229981586337
step: 170, loss: 0.0017472372855991125
step: 180, loss: 0.0022221365943551064
```

```
step: 190, loss: 0.00034926331136375666
step: 200, loss: 0.0001059542628354393
step: 210, loss: 8.18789703771472e-05
step: 220, loss: 0.0002477731031831354
step: 230, loss: 0.0005876274663023651
/usr/local/lib/python3.7/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=0.977728285077951, f1=0.9700996677740864,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.204819121398032e-05
step: 10, loss: 0.00026313064154237751
step: 20, loss: 4.203956268611364e-05
step: 30, loss: 0.0006677695782855153
step: 40, loss: 0.013702243566513062
step: 50, loss: 0.006448269356042147
step: 60, loss: 0.0005060278344899416
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.00017052763723768294
step: 80, loss: 0.0003523677005432546
step: 90, loss: 0.0005758065381087363
step: 100, loss: 0.0005781519575975835
step: 110, loss: 0.00016639060049783438
step: 120, loss: 0.00046141890925355256
step: 130, loss: 5.051098560215905e-05
step: 140, loss: 7.148921577027068e-05
step: 150, loss: 4.488308695727028e-05
step: 160, loss: 0.0001284204190596938
step: 170, loss: 0.0013078056508675218
step: 180, loss: 0.00021758966613560915
step: 190, loss: 0.00023486244026571512
step: 200, loss: 0.00011084515426773578
step: 210, loss: 5.634000990539789e-05
step: 220, loss: 5.0842671043938026e-05
step: 230, loss: 0.013261196203529835
/usr/local/lib/python3.7/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=0.9799107142857142, f1=0.9712389380530975,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011289524263702333
step: 10, loss: 0.0017085670260712504
```

```
step: 20, loss: 4.901372449239716e-05
step: 30, loss: 0.0006173611036501825
step: 40, loss: 5.71494092582725e-05
step: 50, loss: 4.8771067667985335e-05
step: 60, loss: 0.00271209585480392
step: 70, loss: 3.9772276068106294e-05
step: 80, loss: 3.3406700822524726e-05
step: 90, loss: 8.767616236582398e-05
step: 100, loss: 0.000606781744863838
step: 110, loss: 5.6694134400459006e-05
step: 120, loss: 8.169495413312688e-05
step: 130, loss: 0.00010931261931546032
step: 140, loss: 5.865404091309756e-05
step: 150, loss: 0.0016308142803609371
step: 160, loss: 0.02004583366215229
step: 170, loss: 8.737949974602088e-05
step: 180, loss: 5.056299050920643e-05
step: 190, loss: 7.591523171868175e-05
step: 200, loss: 0.004623162094503641
step: 210, loss: 6.54166578897275e-05
step: 220, loss: 0.019419200718402863
step: 230, loss: 6.634195597143844e-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=0.9765886287625419, f1=0.9722530521642618,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011512871424201876
step: 10, loss: 4.6297875087475404e-05
step: 20, loss: 7.695285603404045e-05
step: 30, loss: 0.004625941161066294
step: 40, loss: 0.00012721374514512718
step: 50, loss: 6.726499850628898e-05
step: 60, loss: 8.251853432739154e-05
step: 70, loss: 6.889578071422875e-05
step: 80, loss: 3.364376971148886e-05
step: 90, loss: 3.2542880944674835e-05
step: 100, loss: 0.000340996339218691
step: 110, loss: 0.00012050812802044675
step: 120, loss: 0.00010214278154307976
step: 130, loss: 0.00019587705901358277
step: 140, loss: 0.00013096301699988544
step: 150, loss: 6.455297261709347e-05
step: 160, loss: 5.565610626945272e-05
step: 170, loss: 6.17812984273769e-05
```

```
step: 180, loss: 8.301156049128622e-05
step: 190, loss: 0.0003525045176502317
step: 200, loss: 6.628891424043104e-05
step: 210, loss: 7.599980744998902e-05
step: 220, loss: 7.00959499226883e-05
step: 230, loss: 3.698271757457405e-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=0.980963045912654, f1=0.9722530521642618,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.331437205313705e-05
step: 10, loss: 2.492504245310556e-05
step: 20, loss: 0.01350245252251625
step: 30, loss: 9.190764831146225e-05
step: 40, loss: 3.299278978374787e-05
step: 50, loss: 0.00028177766944281757
step: 60, loss: 3.241173908463679e-05
step: 70, loss: 4.042806176585145e-05
step: 80, loss: 0.00011042927508242428
step: 90, loss: 0.00011328887194395065
step: 100, loss: 0.00045295851305127144
step: 110, loss: 0.0001252306974492967
step: 120, loss: 5.1685303333215415e-05
step: 130, loss: 6.367563764797524e-05
step: 140, loss: 3.108256350969896e-05
step: 150, loss: 6.138469325378537e-05
step: 160, loss: 3.011001172126271e-05
step: 170, loss: 8.592255471739918e-05
step: 180, loss: 5.4664800700265914e-05
step: 190, loss: 0.00019101417274214327
step: 200, loss: 3.965736323152669e-05
step: 210, loss: 0.03323168307542801
step: 220, loss: 5.0571150495670736e-05
step: 230, loss: 2.3610158677911386e-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=0.978865406006674, f1=0.9711751662971175,
best_f1=0.9723145071982282
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.191707219230011e-05
step: 10, loss: 9.467980999033898e-05
```

```
step: 20, loss: 4.9843591114040464e-05
step: 30, loss: 0.00029578135581687093
step: 40, loss: 6.142954953247681e-05
step: 50, loss: 6.0743466747226194e-05
step: 60, loss: 3.0251747375587e-05
step: 70, loss: 0.0009713248000480235
step: 80, loss: 0.0004621469706762582
step: 90, loss: 3.15216529997997e-05
step: 100, loss: 5.224039705353789e-05
step: 110, loss: 8.414156036451459e-05
step: 120, loss: 5.911061816732399e-05
step: 130, loss: 5.236244760453701e-05
step: 140, loss: 0.007275963667780161
step: 150, loss: 7.748482312308624e-05
step: 160, loss: 5.3532497986452654e-05
step: 170, loss: 3.8222631701501086e-05
step: 180, loss: 5.181813321541995e-05
step: 190, loss: 0.00016441001207567751
step: 200, loss: 7.097260822774842e-05
step: 210, loss: 0.0007021464989520609
step: 220, loss: 0.0017331596463918686
step: 230, loss: 0.00020628330821637064
/usr/local/lib/python3.7/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=0.9788182831661093, f1=0.9700996677740864,
best_f1=0.9723145071982282
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
```

```
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 322.22it/s]
load_f1 = 0.9756637168141594
real_f1 = 0.9756637168141594
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, 353.07it/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 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: 0.7888408899307251
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4504043757915497
step: 20, loss: 0.49984490871429443
step: 30, loss: 0.4325622320175171
step: 40, loss: 0.4010685086250305

```
step: 50, loss: 0.23588329553604126
step: 60, loss: 0.18881043791770935
step: 70, loss: 0.23698189854621887
step: 80, loss: 0.12245233356952667
step: 90, loss: 0.06271399557590485
step: 100, loss: 0.2774829864501953
step: 110, loss: 0.08752325177192688
step: 120, loss: 0.05368252843618393
step: 130, loss: 0.015591311268508434
step: 140, loss: 0.1886378824710846
step: 150, loss: 0.029660306870937347
step: 160, loss: 0.1620485931634903
step: 170, loss: 0.1488022804260254
step: 180, loss: 0.08822669088840485
step: 190, loss: 0.018908198922872543
step: 200, loss: 0.09438636153936386
step: 210, loss: 0.08765436708927155
step: 220, loss: 0.08582096546888351
step: 230, loss: 0.09907782822847366
step: 240, loss: 0.0486343689262867
step: 250, loss: 0.04685782268643379
step: 260, loss: 0.041847240179777145
step: 270, loss: 0.007482544519007206
step: 280, loss: 0.10444696247577667
step: 290, loss: 0.07373470067977905
step: 300, loss: 0.09186742454767227
step: 310, loss: 0.08679786324501038
step: 320, loss: 0.053660642355680466
step: 330, loss: 0.04466623067855835
step: 340, loss: 0.06825942546129227
step: 350, loss: 0.04274282231926918
step: 360, loss: 0.0456128753721714
step: 370, loss: 0.14377208054065704
step: 380, loss: 0.17064200341701508
step: 390, loss: 0.02045532315969467
step: 400, loss: 0.015348351560533047
step: 410, loss: 0.025566795840859413
step: 420, loss: 0.01629057712852955
step: 430, loss: 0.06452549993991852
step: 440, loss: 0.0931076779961586
step: 450, loss: 0.026553120464086533
step: 460, loss: 0.17565031349658966
step: 470, loss: 0.18024517595767975
step: 480, loss: 0.43144723773002625
step: 490, loss: 0.03580576553940773
step: 500, loss: 0.0064406683668494225
step: 510, loss: 0.0371549129486084
step: 520, loss: 0.07485076040029526
```

```
step: 530, loss: 0.0311657153069973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9340294257237778, f1=0.9368273280299485,
best_f1=0.9368273280299485
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08518368750810623
step: 10, loss: 0.14125877618789673
step: 20, loss: 0.06298378109931946
step: 30, loss: 0.032434653490781784
step: 40, loss: 0.003027726896107197
step: 50, loss: 0.12663158774375916
step: 60, loss: 0.056749723851680756
step: 70, loss: 0.12498094886541367
step: 80, loss: 0.02109706774353981
step: 90, loss: 0.005899190437048674
step: 100, loss: 0.2511899769306183
step: 110, loss: 0.04556354507803917
step: 120, loss: 0.06915827095508575
step: 130, loss: 0.007848319597542286
step: 140, loss: 0.024800335988402367
step: 150, loss: 0.0727880448102951
step: 160, loss: 0.037252746522426605
step: 170, loss: 0.1484295129776001
step: 180, loss: 0.014795230701565742
step: 190, loss: 0.06352007389068604
step: 200, loss: 0.17366591095924377
step: 210, loss: 0.006547786295413971
step: 220, loss: 0.1606202870607376
step: 230, loss: 0.012790195643901825
step: 240, loss: 0.13388419151306152
step: 250, loss: 0.0068154046311974525
step: 260, loss: 0.012495198287069798
step: 270, loss: 0.09970993548631668
step: 280, loss: 0.16256862878799438
step: 290, loss: 0.0646997019648552
step: 300, loss: 0.0195412989705801
step: 310, loss: 0.06949758529663086
step: 320, loss: 0.11776629090309143
step: 330, loss: 0.044973891228437424
step: 340, loss: 0.01877739280462265
```

```
step: 350, loss: 0.07871881127357483
step: 360, loss: 0.014455248601734638
step: 370, loss: 0.0022210634779185057
step: 380, loss: 0.0715126022696495
step: 390, loss: 0.03648684546351433
step: 400, loss: 0.03012802265584469
step: 410, loss: 0.0008327997056767344
step: 420, loss: 0.10627304762601852
step: 430, loss: 0.017964687198400497
step: 440, loss: 0.016395006328821182
step: 450, loss: 0.01878838613629341
step: 460, loss: 0.24294191598892212
step: 470, loss: 0.02710120752453804
step: 480, loss: 0.1617848128080368
step: 490, loss: 0.02563767321407795
step: 500, loss: 0.022672118619084358
step: 510, loss: 0.059607941657304764
step: 520, loss: 0.093276247382164
step: 530, loss: 0.15434202551841736
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.943500229673863, f1=0.9404216315307058,
best_f1=0.9404216315307058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0055564539507033144
step: 10, loss: 0.05588540807366371
step: 20, loss: 0.21991045773029327
step: 30, loss: 0.15914472937583923
step: 40, loss: 0.0012747903820127249
step: 50, loss: 0.006646858528256416
step: 60, loss: 0.002925141714513302
step: 70, loss: 0.006649219896644354
step: 80, loss: 0.005220633000135422
step: 90, loss: 0.06934582442045212
step: 100, loss: 0.02444782853126526
step: 110, loss: 0.06795301288366318
step: 120, loss: 0.009470603428781033
step: 130, loss: 0.05932148918509483
step: 140, loss: 0.015806300565600395
step: 150, loss: 0.00617886520922184
step: 160, loss: 0.0034596871118992567
```

```
step: 170, loss: 0.00849911943078041
step: 180, loss: 0.07157322764396667
step: 190, loss: 0.07804740220308304
step: 200, loss: 0.0030536246486008167
step: 210, loss: 0.07314334809780121
step: 220, loss: 0.03154249116778374
step: 230, loss: 0.017420515418052673
step: 240, loss: 0.014662428759038448
step: 250, loss: 0.027443422004580498
step: 260, loss: 0.0005521636339835823
step: 270, loss: 0.0008307709940709174
step: 280, loss: 0.007460710126906633
step: 290, loss: 0.09925241768360138
step: 300, loss: 0.06967147439718246
step: 310, loss: 0.14890311658382416
step: 320, loss: 0.137734055519104
step: 330, loss: 0.0016273759538307786
step: 340, loss: 0.003634662600234151
step: 350, loss: 0.07803260535001755
step: 360, loss: 0.009233484975993633
step: 370, loss: 0.0009274314506910741
step: 380, loss: 0.003500369843095541
step: 390, loss: 0.009430599398911
step: 400, loss: 0.004429138731211424
step: 410, loss: 0.007039973046630621
step: 420, loss: 0.03655988350510597
step: 430, loss: 0.04181112349033356
step: 440, loss: 0.08696570247411728
step: 450, loss: 0.02896539494395256
step: 460, loss: 0.11883420497179031
step: 470, loss: 0.008715160191059113
step: 480, loss: 0.0037644889671355486
step: 490, loss: 0.04883740842342377
step: 500, loss: 0.019194861873984337
step: 510, loss: 0.00064739806111902
step: 520, loss: 0.0008620614535175264
step: 530, loss: 0.014834645204246044
/usr/local/lib/python3.7/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=0.9418874941887495, f1=0.9386416861826699,
best_f1=0.9404216315307058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016609880840405822
step: 10, loss: 0.005783040542155504
step: 20, loss: 0.03219512477517128
```

```
step: 30, loss: 0.04149845615029335
step: 40, loss: 0.0029558560345321894
step: 50, loss: 0.006500289775431156
step: 60, loss: 0.0007807380170561373
step: 70, loss: 0.0011942879064008594
step: 80, loss: 0.08602457493543625
step: 90, loss: 0.00048228303785435855
step: 100, loss: 0.003880836069583893
step: 110, loss: 0.006970701273530722
step: 120, loss: 0.0017155497334897518
step: 130, loss: 0.029969656839966774
step: 140, loss: 0.17533451318740845
step: 150, loss: 0.0013221136759966612
step: 160, loss: 0.002676835749298334
step: 170, loss: 0.0019287817412987351
step: 180, loss: 0.0019821380265057087
step: 190, loss: 0.0016554382164031267
step: 200, loss: 0.00025190977612510324
step: 210, loss: 0.019685888662934303
step: 220, loss: 0.0006824146839790046
step: 230, loss: 0.20499582588672638
step: 240, loss: 0.0007186350412666798
step: 250, loss: 0.025177422910928726
step: 260, loss: 0.09465987980365753
step: 270, loss: 0.00574960839003245
step: 280, loss: 0.00856672041118145
step: 290, loss: 0.07734046876430511
step: 300, loss: 0.0015957494033500552
step: 310, loss: 0.00550222210586071
step: 320, loss: 0.026160389184951782
step: 330, loss: 0.049972053617239
step: 340, loss: 0.0012278191279619932
step: 350, loss: 0.00031286923331208527
step: 360, loss: 0.016465911641716957
step: 370, loss: 0.03923184052109718
step: 380, loss: 0.004724208265542984
step: 390, loss: 0.02885877899825573
step: 400, loss: 0.01227312907576561
step: 410, loss: 0.003991239238530397
step: 420, loss: 0.0021229588892310858
step: 430, loss: 0.0013091234723106027
step: 440, loss: 0.03476708382368088
step: 450, loss: 0.027648719027638435
step: 460, loss: 0.0010837777517735958
step: 470, loss: 0.0016496196622028947
step: 480, loss: 0.034747347235679626
step: 490, loss: 0.019908368587493896
step: 500, loss: 0.016928572207689285
```

```
step: 510, loss: 0.011819603852927685
step: 520, loss: 0.00041115510975942016
step: 530, loss: 0.012813953682780266
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9441624365482235, f1=0.9388888888888889,
best_f1=0.9388888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002533664694055915
step: 10, loss: 0.001966190757229924
step: 20, loss: 0.00060202059103176
step: 30, loss: 0.0008045788272283971
step: 40, loss: 0.009676925837993622
step: 50, loss: 0.0155516117811203
step: 60, loss: 0.003667603014037013
step: 70, loss: 0.00016919443442020565
step: 80, loss: 0.0003914835106115788
step: 90, loss: 0.0021838475950062275
step: 100, loss: 0.0002713843423407525
step: 110, loss: 0.01089190412312746
step: 120, loss: 0.020900998264551163
step: 130, loss: 0.0016232893103733659
step: 140, loss: 0.0006106223445385695
step: 150, loss: 0.005298634525388479
step: 160, loss: 0.08578187972307205
step: 170, loss: 0.041586995124816895
step: 180, loss: 0.0070055206306278706
step: 190, loss: 0.0016213060589507222
step: 200, loss: 0.0003536323201842606
step: 210, loss: 0.0004529940488282591
step: 220, loss: 0.0001068890342139639
step: 230, loss: 0.0024853101931512356
step: 240, loss: 0.0015114954439923167
step: 250, loss: 0.0007396640721708536
step: 260, loss: 0.0873834416270256
step: 270, loss: 0.03955203667283058
step: 280, loss: 0.03597480431199074
step: 290, loss: 0.0005545421154238284
step: 300, loss: 0.018128393217921257
step: 310, loss: 0.028745893388986588
step: 320, loss: 0.007923673838376999
```

```
step: 330, loss: 0.0005317724426276982
step: 340, loss: 0.010055728256702423
step: 350, loss: 0.0001638872636249289
step: 360, loss: 0.00657209288328886
step: 370, loss: 0.000345598062267527
step: 380, loss: 0.010432299226522446
step: 390, loss: 0.004597052466124296
step: 400, loss: 0.0026783542707562447
step: 410, loss: 0.002129461383447051
step: 420, loss: 0.004281855653971434
step: 430, loss: 0.0023113014176487923
step: 440, loss: 0.0011634811526164412
step: 450, loss: 0.026799382641911507
step: 460, loss: 0.0003795235534198582
step: 470, loss: 0.0011299289762973785
step: 480, loss: 0.029554175212979317
step: 490, loss: 0.003283401019871235
step: 500, loss: 0.022434063255786896
step: 510, loss: 0.0013676552334800363
step: 520, loss: 0.0001810885441955179
step: 530, loss: 0.015843473374843597
/usr/local/lib/python3.7/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=0.9433962264150944, f1=0.9326375711574952,
best_f1=0.9388888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003727188741322607
step: 10, loss: 0.0012876065447926521
step: 20, loss: 0.000650838715955615
step: 30, loss: 0.010257354937493801
step: 40, loss: 0.00043089388054795563
step: 50, loss: 0.0008299691253341734
step: 60, loss: 0.0005697807646356523
step: 70, loss: 0.0037848660722374916
step: 80, loss: 0.0005575203103944659
step: 90, loss: 0.0014854512410238385
step: 100, loss: 0.018703928217291832
step: 110, loss: 0.0008097616373561323
step: 120, loss: 0.0002476142253726721
step: 130, loss: 7.747980998829007e-05
step: 140, loss: 0.0009551577968522906
step: 150, loss: 0.000546040479093790
step: 160, loss: 7.67305537010543e-05
step: 170, loss: 0.00010424535867059603
step: 180, loss: 0.00011134459055028856
```

```
step: 190, loss: 0.12673614919185638
step: 200, loss: 0.00025853878469206393
step: 210, loss: 0.00018706951232161373
step: 220, loss: 0.00024568181834183633
step: 230, loss: 0.0001452276628697291
step: 240, loss: 0.00025842650211416185
step: 250, loss: 0.00014569441555067897
step: 260, loss: 0.00026097751106135547
step: 270, loss: 0.00031487317755818367
step: 280, loss: 0.0010216577211394906
step: 290, loss: 0.00023078755475580692
step: 300, loss: 0.0003751321928575635
step: 310, loss: 0.0005425630370154977
step: 320, loss: 0.011898989789187908
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 330, loss: 0.07887712866067886
step: 340, loss: 0.0014411587035283446
step: 350, loss: 0.018936993554234505
step: 360, loss: 8.601289300713688e-05
step: 370, loss: 0.0010314162354916334
step: 380, loss: 0.011103627271950245
step: 390, loss: 0.013037689961493015
step: 400, loss: 0.00028005559579469264
step: 410, loss: 0.0002551275829318911
step: 420, loss: 0.000449318322353065
step: 430, loss: 0.00023909969604574144
step: 440, loss: 0.12450717389583588
step: 450, loss: 0.0017024873523041606
step: 460, loss: 0.0007371741230599582
step: 470, loss: 0.10964687913656235
step: 480, loss: 0.0062161837704479694
step: 490, loss: 0.00015658210031688213
step: 500, loss: 0.0016727831680327654
step: 510, loss: 0.0001755681005306542
step: 520, loss: 0.0030085279140621424
step: 530, loss: 0.011319171637296677
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9458955223880597, f1=0.9367088607594937,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.002436110284179449
step: 10, loss: 0.0008680531173013151
step: 20, loss: 0.009280635043978691
step: 30, loss: 0.001269992091692984
step: 40, loss: 0.022498030215501785
step: 50, loss: 0.0153376255184412
step: 60, loss: 0.00013939954806119204
step: 70, loss: 0.002331587253138423
step: 80, loss: 0.001530132838524878
step: 90, loss: 0.00011652892862912267
step: 100, loss: 0.0017831713194027543
step: 110, loss: 0.0033897790126502514
step: 120, loss: 0.00018921667651738971
step: 130, loss: 0.0030086799524724483
step: 140, loss: 0.002409039530903101
step: 150, loss: 0.00014860121882520616
step: 160, loss: 0.001942358328960836
step: 170, loss: 0.04635614529252052
step: 180, loss: 0.00011013129551429302
step: 190, loss: 8.706306107342243e-05
step: 200, loss: 0.00014362091314978898
step: 210, loss: 0.0001271418295800686
step: 220, loss: 7.566650310764089e-05
step: 230, loss: 0.01548828650265932
step: 240, loss: 0.0018354121129959822
step: 250, loss: 9.012024383991957e-05
step: 260, loss: 0.0003813765069935471
step: 270, loss: 4.6301563997985795e-05
step: 280, loss: 0.0038422911893576384
step: 290, loss: 0.004153990186750889
step: 300, loss: 0.00011902310507139191
step: 310, loss: 0.0001846009836299345
step: 320, loss: 0.02482868731021881
step: 330, loss: 0.00010324579488951713
step: 340, loss: 0.004459071438759565
step: 350, loss: 0.003908935002982616
step: 360, loss: 0.003962937742471695
step: 370, loss: 0.00018945385818369687
step: 380, loss: 0.0005354215390980244
step: 390, loss: 0.0011751970741897821
step: 400, loss: 5.920007606619038e-05
step: 410, loss: 0.0005250925314612687
step: 420, loss: 8.550490747438744e-05
step: 430, loss: 8.389871072722599e-05
step: 440, loss: 6.491304520750418e-05
step: 450, loss: 8.986095053842291e-05
step: 460, loss: 0.0007953911554068327
step: 470, loss: 0.0170725230127573
```

```
step: 480, loss: 0.0010491044959053397
step: 490, loss: 0.004382276441901922
step: 500, loss: 5.9222780691925436e-05
step: 510, loss: 0.005365576595067978
step: 520, loss: 0.002892390126362443
step: 530, loss: 0.0001220265548909083
/usr/local/lib/python3.7/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=0.9438515081206497, f1=0.9349442379182156,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013577513163909316
step: 10, loss: 0.010665182024240494
step: 20, loss: 0.035645436495542526
step: 30, loss: 0.0004566300194710493
step: 40, loss: 0.00011089895269833505
step: 50, loss: 0.0003837333351839334
step: 60, loss: 0.004517309833317995
step: 70, loss: 0.0001626737939659506
step: 80, loss: 0.00033379753585904837
step: 90, loss: 8.432596951024607e-05
step: 100, loss: 0.000343110179528594
step: 110, loss: 0.0001526197011116892
step: 120, loss: 0.0009516640566289425
step: 130, loss: 0.00016757127013988793
step: 140, loss: 3.150327393086627e-05
step: 150, loss: 0.00029663709574379027
step: 160, loss: 6.491399835795164e-05
step: 170, loss: 6.123719504103065e-05
step: 180, loss: 0.0001474244345445186
step: 190, loss: 0.00014812186418566853
step: 200, loss: 0.0010823969496414065
step: 210, loss: 0.0005755048478022218
step: 220, loss: 0.001446122769266367
step: 230, loss: 0.007461662869900465
step: 240, loss: 6.28351845080033e-05
step: 250, loss: 0.020040320232510567
step: 260, loss: 0.0006180984200909734
step: 270, loss: 3.25206492561847e-05
step: 280, loss: 0.0038792130071669817
step: 290, loss: 0.0008610779186710715
step: 300, loss: 3.503781772451475e-05
step: 310, loss: 0.007630478125065565
step: 320, loss: 9.38312805374153e-05
step: 330, loss: 0.0724642425775528
```

```
step: 340, loss: 0.004308800678700209
step: 350, loss: 0.0014901243848726153
step: 360, loss: 0.0004009223193861544
step: 370, loss: 0.0006032360251992941
step: 380, loss: 0.00024021438730414957
step: 390, loss: 7.636036752955988e-05
step: 400, loss: 0.05831896513700485
step: 410, loss: 0.028053628280758858
step: 420, loss: 0.1210426315665245
step: 430, loss: 0.00036119919968768954
step: 440, loss: 0.012748770415782928
step: 450, loss: 0.001375738182105124
step: 460, loss: 0.0002318512852070853
step: 470, loss: 0.00012000483548035845
step: 480, loss: 0.007877795957028866
step: 490, loss: 0.00018501936574466527
step: 500, loss: 0.0006022467277944088
step: 510, loss: 3.9326365367742255e-05
step: 520, loss: 0.00016930171113926917
step: 530, loss: 0.0005726073868572712
/usr/local/lib/python3.7/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=0.9350282485875707, f1=0.9268755935422602,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012020643043797463
step: 10, loss: 4.5974928070791066e-05
step: 20, loss: 7.948146958369762e-05
step: 30, loss: 0.00258851470425725
step: 40, loss: 0.0005758590996265411
step: 50, loss: 0.0008598514250479639
step: 60, loss: 9.056131239049137e-05
step: 70, loss: 0.14239813387393951
step: 80, loss: 0.0001362164766760543
step: 90, loss: 0.00020931589824613184
step: 100, loss: 0.00023785411030985415
step: 110, loss: 0.00024656965979374945
step: 120, loss: 4.8891230107983574e-05
step: 130, loss: 0.009902358055114746
step: 140, loss: 0.00047738576540723443
step: 150, loss: 3.8379446777980775e-05
step: 160, loss: 0.00034124994999729097
step: 170, loss: 3.938207737519406e-05
step: 180, loss: 5.002223770134151e-05
step: 190, loss: 0.0002053480566246435
```

```
step: 200, loss: 0.0004928524140268564
step: 210, loss: 6.800445407861844e-05
step: 220, loss: 0.0003257211938034743
step: 230, loss: 0.0022965397220104933
step: 240, loss: 0.0003693445469252765
step: 250, loss: 0.002702901139855385
step: 260, loss: 0.20527884364128113
step: 270, loss: 0.0012366857845336199
step: 280, loss: 8.355415775440633e-05
step: 290, loss: 9.071581007447094e-05
step: 300, loss: 5.795766628580168e-05
step: 310, loss: 5.247112494544126e-05
step: 320, loss: 0.0003638010530266911
step: 330, loss: 0.0003437835839577019
step: 340, loss: 3.751167605514638e-05
step: 350, loss: 3.9378417568514124e-05
step: 360, loss: 0.041051022708415985
step: 370, loss: 6.128360837465152e-05
step: 380, loss: 3.986590672866441e-05
step: 390, loss: 0.00034242449328303337
step: 400, loss: 0.00167095719370991
step: 410, loss: 8.384958346141502e-05
step: 420, loss: 0.01112013403326273
step: 430, loss: 3.109793033218011e-05
step: 440, loss: 0.00044199475087225437
step: 450, loss: 0.0001298493007197976
step: 460, loss: 0.00025104504311457276
step: 470, loss: 0.00010335533443139866
step: 480, loss: 0.0005704876384697855
step: 490, loss: 8.500379044562578e-05
step: 500, loss: 0.01778373122215271
step: 510, loss: 0.0002232748083770275
step: 520, loss: 6.256993947317824e-05
step: 530, loss: 0.009107168763875961
/usr/local/lib/python3.7/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=0.9424964936886395, f1=0.9311320754716981,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003333355998620391
step: 10, loss: 7.283256127266213e-05
step: 20, loss: 0.0006815530359745026
step: 30, loss: 0.0010886816307902336
step: 40, loss: 0.0008932246710173786
step: 50, loss: 0.0007974962936714292
```

```
step: 60, loss: 0.00035801585181616247
step: 70, loss: 0.0006464255275204778
step: 80, loss: 0.0010109145659953356
step: 90, loss: 0.00023910895106382668
step: 100, loss: 0.00011558199184946716
step: 110, loss: 0.00091811508173123
step: 120, loss: 0.00013388789375312626
step: 130, loss: 5.187817077967338e-05
step: 140, loss: 7.260466372827068e-05
step: 150, loss: 6.126216612756252e-05
step: 160, loss: 0.00013354237307794392
step: 170, loss: 0.0006360894185490906
step: 180, loss: 0.009098718874156475
step: 190, loss: 0.0001493830350227654
step: 200, loss: 7.987737626535818e-05
step: 210, loss: 4.0350652852794155e-05
step: 220, loss: 3.885341720888391e-05
step: 230, loss: 0.0021865232847630978
step: 240, loss: 4.1509436414344236e-05
step: 250, loss: 0.0020991608034819365
step: 260, loss: 0.07459142059087753
step: 270, loss: 0.00016720335406716913
step: 280, loss: 0.00023747239902149886
step: 290, loss: 0.00011999468551948667
step: 300, loss: 0.004063013009727001
step: 310, loss: 0.0011914483038708568
step: 320, loss: 0.0006707340944558382
step: 330, loss: 0.000523876107763499
step: 340, loss: 0.00011154902929556556
step: 350, loss: 4.698099655797705e-05
step: 360, loss: 0.003309119027107954
step: 370, loss: 0.005475231911987066
step: 380, loss: 0.0320127010345459
step: 390, loss: 5.782551670563407e-05
step: 400, loss: 9.8767428426072e-05
step: 410, loss: 0.00024383327399846166
step: 420, loss: 0.00024104789190459996
step: 430, loss: 6.951280374778435e-05
step: 440, loss: 3.765021392609924e-05
step: 450, loss: 0.005631780251860619
step: 460, loss: 0.0014261923497542739
step: 470, loss: 3.736713188118301e-05
step: 480, loss: 0.0006939079030416906
step: 490, loss: 0.06401350349187851
step: 500, loss: 0.025481870397925377
step: 510, loss: 0.0006543413619510829
step: 520, loss: 0.0002865529677364975
step: 530, loss: 2.4947725250967778e-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=0.9452690166975881, f1=0.9333958724202628,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00044965604320168495
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.0010874951258301735
step: 20, loss: 0.0005286551895551383
step: 30, loss: 0.002152013126760721
step: 40, loss: 0.016651052981615067
step: 50, loss: 0.00015051974332891405
step: 60, loss: 0.00010294989624526352
step: 70, loss: 5.365649121813476e-05
step: 80, loss: 0.00011618674761848524
step: 90, loss: 0.0015609218971803784
step: 100, loss: 0.00027247663820162416
step: 110, loss: 0.00012629985576495528
step: 120, loss: 0.000985948834568262
step: 130, loss: 0.0012090152595192194
step: 140, loss: 4.357647412689403e-05
step: 150, loss: 0.0001215269512613304
step: 160, loss: 0.00037903175689280033
step: 170, loss: 0.0006565276416949928
step: 180, loss: 7.566613203380257e-05
step: 190, loss: 0.07959482073783875
step: 200, loss: 0.0039895265363156796
step: 210, loss: 9.494332334725186e-05
step: 220, loss: 0.00020681091700680554
step: 230, loss: 0.00021122506586834788
step: 240, loss: 0.0002711534034460783
step: 250, loss: 4.432651985553093e-05
step: 260, loss: 0.0008958937251009047
step: 270, loss: 0.0008738025208003819
step: 280, loss: 0.00014874694170430303
step: 290, loss: 0.0006768118473701179
step: 300, loss: 0.011605623178184032
step: 310, loss: 3.5690132790477946e-05
step: 320, loss: 0.00970497727394104
step: 330, loss: 0.0010303169256076217
step: 340, loss: 3.716899300343357e-05
step: 350, loss: 0.00048579490976408124
step: 360, loss: 6.9961700122874901e-05
step: 370, loss: 2.9775423172395676e-05
step: 380, loss: 4.965329571859911e-05
```

```
step: 390, loss: 0.026783453300595284
step: 400, loss: 3.995148654212244e-05
step: 410, loss: 0.00010009047400671989
step: 420, loss: 0.003729657270014286
step: 430, loss: 0.004771714564412832
step: 440, loss: 7.292685040738434e-05
step: 450, loss: 3.407055191928521e-05
step: 460, loss: 0.00644173426553607
step: 470, loss: 0.006293143145740032
step: 480, loss: 0.0002685078652575612
step: 490, loss: 0.1413855105638504
step: 500, loss: 0.00017475432832725346
step: 510, loss: 6.293600017670542e-05
step: 520, loss: 0.00027439792756922543
step: 530, loss: 0.00014711034600622952
/usr/local/lib/python3.7/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=0.9433962264150944, f1=0.9307400379506643,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020595172827597708
step: 10, loss: 0.0004590170574374497
step: 20, loss: 0.0002611610689200461
step: 30, loss: 9.350429172627628e-05
step: 40, loss: 5.564743696595542e-05
step: 50, loss: 0.0006651714211329818
step: 60, loss: 0.009676271118223667
step: 70, loss: 0.0035048355348408222
step: 80, loss: 0.002298020990565419
step: 90, loss: 9.75749280769378e-05
step: 100, loss: 0.003936399705708027
step: 110, loss: 0.0002991647052112967
step: 120, loss: 0.00019749492639675736
step: 130, loss: 8.131725917337462e-05
step: 140, loss: 5.843200415256433e-05
step: 150, loss: 0.0023043814580887556
step: 160, loss: 0.00011889497545780614
step: 170, loss: 0.00011604710016399622
step: 180, loss: 0.00016831118607616052
step: 190, loss: 4.7322348109446466e-05
step: 200, loss: 7.919142808532342e-05
step: 210, loss: 0.0006417600670829415
step: 220, loss: 2.8058171665179543e-05
step: 230, loss: 0.00401280727237463
step: 240, loss: 0.00011981749412370846
```

```
step: 250, loss: 4.074458411196247e-05
step: 260, loss: 0.00011594210081966594
step: 270, loss: 9.744644921738654e-05
step: 280, loss: 5.119858542457223e-05
step: 290, loss: 0.02785416692495346
step: 300, loss: 8.390657603740692e-05
step: 310, loss: 4.5806194975739345e-05
step: 320, loss: 3.332169580971822e-05
step: 330, loss: 3.974984065280296e-05
step: 340, loss: 0.0002665857900865376
step: 350, loss: 0.003439798252657056
step: 360, loss: 0.0003748380404431373
step: 370, loss: 3.336622830829583e-05
step: 380, loss: 3.647650373750366e-05
step: 390, loss: 6.917122664162889e-05
step: 400, loss: 3.446817572694272e-05
step: 410, loss: 0.013301355764269829
step: 420, loss: 0.0003220572543796152
step: 430, loss: 3.420433131395839e-05
step: 440, loss: 5.309931293595582e-05
step: 450, loss: 0.0005336682661436498
step: 460, loss: 3.81560530513525e-05
step: 470, loss: 4.213606734992936e-05
step: 480, loss: 0.00014745749649591744
step: 490, loss: 4.316574631957337e-05
step: 500, loss: 0.0025137262418866158
step: 510, loss: 4.426162922754884e-05
step: 520, loss: 0.01624329946935177
step: 530, loss: 0.0002308513649040833
/usr/local/lib/python3.7/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=0.9450045829514206, f1=0.9355432780847146,
best_f1=0.9367088607594937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0029143965803086758
step: 10, loss: 0.00010256233508698642
step: 20, loss: 0.026018425822257996
step: 30, loss: 0.0010929340496659279
step: 40, loss: 4.783995609614067e-05
step: 50, loss: 0.006248122081160545
step: 60, loss: 4.761587115353905e-05
step: 70, loss: 3.7857596907997504e-05
step: 80, loss: 3.0341510864673182e-05
step: 90, loss: 0.000928056426346302
step: 100, loss: 4.485321551328525e-05
```

```
step: 110, loss: 4.714267925010063e-05
step: 120, loss: 7.691326027270406e-05
step: 130, loss: 0.0029844618402421474
step: 140, loss: 0.0035615640226751566
step: 150, loss: 3.507586370687932e-05
step: 160, loss: 6.817909888923168e-05
step: 170, loss: 9.109706297749653e-05
step: 180, loss: 2.9998545869602822e-05
step: 190, loss: 9.676544141257182e-05
step: 200, loss: 0.0022481633350253105
step: 210, loss: 0.000498484296258539
step: 220, loss: 0.0007778875296935439
step: 230, loss: 2.9342054403969087e-05
step: 240, loss: 4.954118776367977e-05
step: 250, loss: 0.07766655832529068
step: 260, loss: 0.00014592809020541608
step: 270, loss: 0.0002101813443005085
step: 280, loss: 0.0010152359027415514
step: 290, loss: 8.203698962461203e-05
step: 300, loss: 7.735453982604668e-05
step: 310, loss: 0.002873932244256139
step: 320, loss: 7.99679328338243e-05
step: 330, loss: 0.0003465153858996928
step: 340, loss: 3.675865809782408e-05
step: 350, loss: 0.030962355434894562
step: 360, loss: 3.0311020964290947e-05
step: 370, loss: 3.657607521745376e-05
step: 380, loss: 0.0015679440693929791
step: 390, loss: 2.8911261324537918e-05
step: 400, loss: 3.663900861283764e-05
step: 410, loss: 4.09752938139718e-05
step: 420, loss: 4.444172736839391e-05
step: 430, loss: 7.252647628774866e-05
step: 440, loss: 0.00013216593652032316
step: 450, loss: 0.0009259369107894599
step: 460, loss: 2.2693830032949336e-05
step: 470, loss: 0.02147805690765381
step: 480, loss: 0.0015185987576842308
step: 490, loss: 7.132026803446934e-05
step: 500, loss: 0.0017842997331172228
step: 510, loss: 0.0013944071251899004
step: 520, loss: 3.0277602490969002e-05
step: 530, loss: 5.6476077588740736e-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=0.9463955637707948, f1=0.9397590361445785,
best_f1=0.9397590361445785
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.115361611591652e-05
step: 10, loss: 0.00017701269825920463
step: 20, loss: 2.28243679885054e-05
step: 30, loss: 4.313311364967376e-05
step: 40, loss: 4.704284583567642e-05
step: 50, loss: 9.58698001340963e-05
step: 60, loss: 3.0061573852435686e-05
step: 70, loss: 2.392323403910268e-05
step: 80, loss: 0.00011371784785296768
step: 90, loss: 2.0529678295133635e-05
step: 100, loss: 0.005431467201560736
step: 110, loss: 0.00024347170256078243
step: 120, loss: 1.608925413165707e-05
step: 130, loss: 0.00021962983009871095
step: 140, loss: 4.7356359573313966e-05
step: 150, loss: 6.970685353735462e-05
step: 160, loss: 0.0037849610671401024
step: 170, loss: 0.00018457842816133052
step: 180, loss: 3.071012542932294e-05
step: 190, loss: 0.0001050419159582816
step: 200, loss: 0.00019791733939320064
step: 210, loss: 0.0002703883801586926
step: 220, loss: 3.92058354918845e-05
step: 230, loss: 0.0012324880808591843
step: 240, loss: 0.00017429041326977313
step: 250, loss: 4.477091715671122e-05
step: 260, loss: 0.00010235162335447967
step: 270, loss: 0.0025804340839385986
step: 280, loss: 4.7034991439431906e-05
step: 290, loss: 0.0007375759305432439
step: 300, loss: 8.128426998155192e-05
step: 310, loss: 3.483169348328374e-05
step: 320, loss: 0.00044616361265070736
step: 330, loss: 4.1994295315817e-05
step: 340, loss: 3.117963933618739e-05
step: 350, loss: 0.00021020264830440283
step: 360, loss: 0.0011443227995187044
step: 370, loss: 3.45452172041405e-05
step: 380, loss: 2.5200693926308304e-05
step: 390, loss: 2.8825268600485288e-05
step: 400, loss: 4.408410677569918e-05
```

```
step: 410, loss: 2.371457412664313e-05
step: 420, loss: 2.099538323818706e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 430, loss: 5.6045995734166354e-05
step: 440, loss: 1.4792974980082363e-05
step: 450, loss: 4.25534053647425e-05
step: 460, loss: 0.00015593816351611167
step: 470, loss: 0.00014699033636134118
step: 480, loss: 0.0001234445080626756
step: 490, loss: 2.7804164346889593e-05
step: 500, loss: 3.311967884656042e-05
step: 510, loss: 6.106817454565316e-05
step: 520, loss: 3.7295765650924295e-05
step: 530, loss: 3.302978439023718e-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=0.9444958371877892, f1=0.9327146171693736,
best_f1=0.9397590361445785
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.3587837858940475e-05
step: 10, loss: 0.0012275767512619495
step: 20, loss: 4.943479143548757e-05
step: 30, loss: 0.00010450738773215562
step: 40, loss: 1.4085170732869301e-05
step: 50, loss: 0.000502164417412132
step: 60, loss: 0.0002707515668589622
step: 70, loss: 8.694344433024526e-05
step: 80, loss: 3.426986950216815e-05
step: 90, loss: 3.6522847949527204e-05
step: 100, loss: 3.0981351301306860e-05
step: 110, loss: 0.008212885819375515
step: 120, loss: 0.0014823238598182797
step: 130, loss: 1.53665678226389e-05
step: 140, loss: 1.9054405129281804e-05
step: 150, loss: 6.050922820577398e-05
step: 160, loss: 3.2262854801956564e-05
step: 170, loss: 3.653165913419798e-05
step: 180, loss: 2.2183479813975282e-05
step: 190, loss: 0.00010746666521299963
step: 200, loss: 6.696662603644654e-05
step: 210, loss: 3.5303994081914425e-05
step: 220, loss: 1.3414639397524297e-05
step: 230, loss: 2.5796731279115193e-05
step: 240, loss: 3.8173871871549636e-05
step: 250, loss: 7.290105713764206e-05
```

```
step: 260, loss: 4.865641676587984e-05
step: 270, loss: 0.00012720815720967948
step: 280, loss: 0.00029715863638557494
step: 290, loss: 0.007110943552106619
step: 300, loss: 0.00011131894279969856
step: 310, loss: 0.00011991860810667276
step: 320, loss: 4.803468254976906e-05
step: 330, loss: 4.711225483333692e-05
step: 340, loss: 4.1636991227278486e-05
step: 350, loss: 4.5466287701856345e-05
step: 360, loss: 0.00015034610987640917
step: 370, loss: 6.007348929415457e-05
step: 380, loss: 2.1311776436050422e-05
step: 390, loss: 2.2880012693349272e-05
step: 400, loss: 4.3619194912025705e-05
step: 410, loss: 0.000274116377113387
step: 420, loss: 4.922435982734896e-05
step: 430, loss: 3.86956162401475e-05
step: 440, loss: 0.00021708181884605438
step: 450, loss: 0.0014256108552217484
step: 460, loss: 2.162075361411553e-05
step: 470, loss: 4.22762350353878e-05
step: 480, loss: 2.1710324290324934e-05
step: 490, loss: 4.043818626087159e-05
step: 500, loss: 0.0005367069388739765
step: 510, loss: 0.0009509872179478407
step: 520, loss: 5.3779385780217126e-05
step: 530, loss: 8.376777986995876e-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 15: dev_f1=0.9464450600184672, f1=0.9388888888888889,
best_f1=0.9388888888888889
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
```

```
loss_scale              : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                 : True
opt_level               : O2
cast_model_type         : torch.float16
patch_torch_functions   : False
keep_batchnorm_fp32     : True
master_weights          : True
loss_scale              : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:16, 353.64it/s]
load_f1 = 0.9434843531060252
real_f1 = 0.9415614773258532
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, 356.33it/s]
```

4 DITTO

4.1 DITTO STRUCTURED

4.1.1 Beer - Running the matcher

```python
# 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 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: 0.8634351491928101
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=0.3333333333333333, f1=0.11764705882352941,
best_f1=0.11764705882352941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37270042300224304
/usr/local/lib/python3.7/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=0.288659793814433, f1=0.2857142857142857,
best_f1=0.11764705882352941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.31200364232063293
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.48484848484848486, f1=0.3888888888888889,
best_f1=0.3888888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.35872986912727356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.5098039215686275, f1=0.4150943396226415,
best_f1=0.4150943396226415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.257152795791626
/usr/local/lib/python3.7/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=0.47058823529411764, f1=0.5, best_f1=0.4150943396226415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.290949285030365
/usr/local/lib/python3.7/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=0.4705882352941177, f1=0.423076923076923,
best_f1=0.4150943396226415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19756528735160828
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5128205128205129, f1=0.4545454545454545,
best_f1=0.4545454545454545
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32338422536849976
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.64, f1=0.4827586206896552, best_f1=0.4827586206896552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14132505655288696
/usr/local/lib/python3.7/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=0.6206896551724138, f1=0.5142857142857143,
best_f1=0.4827586206896552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23538315296173096
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.689655172413793, 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: 0.20698945224285126
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6956521739130435, f1=0.4827586206896552,
best_f1=0.4827586206896552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16966257989406586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.75, f1=0.4827586206896552, best_f1=0.4827586206896552
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11336242407560349
/usr/local/lib/python3.7/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=0.7333333333333334, f1=0.5714285714285714,
best_f1=0.4827586206896552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23579095304012299
/usr/local/lib/python3.7/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=0.7333333333333334, f1=0.5454545454545454,
best_f1=0.4827586206896552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16356965899467468
/usr/local/lib/python3.7/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=0.7333333333333334, f1=0.5454545454545454,
best_f1=0.4827586206896552
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.")
91it [00:00, 141625.85it/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 = 0.6666666666666667
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, 383.21it/s]
```

4.1.2 DBLP-ACM - Running the matcher

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


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

     # Step 3: run the matcher
     !CUDA_VISIBLE_DEVICES=0 python matcher.py \
       --task Structured/DBLP-ACM \
```

```
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 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: 0.8056520819664001
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.46708232164382935
step: 20, loss: 0.6094183921813965

```
step: 30, loss: 0.4535389542579651
step: 40, loss: 0.2739093005657196
step: 50, loss: 0.08671186119318008
step: 60, loss: 0.07209192216396332
step: 70, loss: 0.1466364562511444
step: 80, loss: 0.17276722192764282
step: 90, loss: 0.08123919367790222
step: 100, loss: 0.30120664834976196
step: 110, loss: 0.14035673439502716
step: 120, loss: 0.022027984261512756
step: 130, loss: 0.007864844053983688
step: 140, loss: 0.008595661260187626
step: 150, loss: 0.08903941512107849
step: 160, loss: 0.14281703531742096
step: 170, loss: 0.030738718807697296
step: 180, loss: 0.006748225074261427
step: 190, loss: 0.09328649193048477
step: 200, loss: 0.01144753210246563
step: 210, loss: 0.10504581779241562
step: 220, loss: 0.005838553886860609
step: 230, loss: 0.0036478289403021336
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9796380090497738, f1=0.9796380090497738,
best_f1=0.9796380090497738
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14872680604457855
step: 10, loss: 0.0019273965153843164
step: 20, loss: 0.0037415074184536934
step: 30, loss: 0.0023456579074263573
step: 40, loss: 0.046435609459877014
step: 50, loss: 0.0033530823420733213
step: 60, loss: 0.0021764966659247875
step: 70, loss: 0.004553234204649925
step: 80, loss: 0.006417127791792154
step: 90, loss: 0.006722652353346348
step: 100, loss: 0.02827932871878147
step: 110, loss: 0.033762816339731216
step: 120, loss: 0.0011414604960009456
step: 130, loss: 0.007799738086760044
step: 140, loss: 0.07737620919942856
```

```
step: 150, loss: 0.005947126541286707
step: 160, loss: 0.00436435779556632
step: 170, loss: 0.009030819870531559
step: 180, loss: 0.001979569438844919
step: 190, loss: 0.0991039127111435
step: 200, loss: 0.008495545946061611
step: 210, loss: 0.141254723072052
step: 220, loss: 0.0046242144489012957
step: 230, loss: 0.08761497586965561
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9831271091113611, f1=0.9819004524886877,
best_f1=0.9819004524886877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08332382887601852
step: 10, loss: 0.030700335279107094
step: 20, loss: 0.0063739861361682415
step: 30, loss: 0.030336106196045876
step: 40, loss: 0.07222913950681686
step: 50, loss: 0.0032672234810888767
step: 60, loss: 0.009962782263755798
step: 70, loss: 0.004020977765321732
step: 80, loss: 0.04597293958067894
step: 90, loss: 0.011429399251937866
step: 100, loss: 0.004514267202466726
step: 110, loss: 0.0050031510181725025
step: 120, loss: 0.0032113397028297186
step: 130, loss: 0.0012131805997341871
step: 140, loss: 0.001477265846915543
step: 150, loss: 0.0010891336714848876
step: 160, loss: 0.0007344837067648768
step: 170, loss: 0.003991982433944941
step: 180, loss: 0.0014963747235015035
step: 190, loss: 0.0023072639014571905
step: 200, loss: 0.02556055411696434
step: 210, loss: 0.02045254595577717
step: 220, loss: 0.0005281638586893678
step: 230, loss: 0.032563354820013046
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9887640449438202, f1=0.9898305084745763,
best_f1=0.9898305084745763
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008263663039542735
step: 10, loss: 0.003557311836630106
step: 20, loss: 0.00091046869056629933
step: 30, loss: 0.0005258747260086238
step: 40, loss: 0.0008928063907660544
step: 50, loss: 0.0016419633757323027
step: 60, loss: 0.00042694562580043647
step: 70, loss: 0.02665209211409092
step: 80, loss: 0.06876413524150848
step: 90, loss: 0.0020045372657477856
step: 100, loss: 0.05378061905503273
step: 110, loss: 0.028006182983517647
step: 120, loss: 0.009739368222653866
step: 130, loss: 0.00328094232827425
step: 140, loss: 0.00497082294896245
step: 150, loss: 0.045187849551439285
step: 160, loss: 0.00038722591125406325
step: 170, loss: 0.14465582370758057
step: 180, loss: 0.03849712386727333
step: 190, loss: 0.019099989905953407
step: 200, loss: 0.0003848435590043664
step: 210, loss: 0.0012398201506584883
step: 220, loss: 0.0007669642218388617
step: 230, loss: 0.00016159190272446722
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9887892376681614, f1=0.9854423292273236,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005494328215718269
step: 10, loss: 0.00017271877732127905
step: 20, loss: 0.0002448874292895198

```
step: 30, loss: 0.003754980396479368
step: 40, loss: 0.0001666703901719302
step: 50, loss: 0.039677660912275314
step: 60, loss: 0.0004331759409978986
step: 70, loss: 0.00013033434515818954
step: 80, loss: 0.032555483281612396
step: 90, loss: 0.0016771652735769749
step: 100, loss: 0.0018104406772181392
step: 110, loss: 0.024813294410705566
step: 120, loss: 0.04366900399327278
step: 130, loss: 0.047696154564619064
step: 140, loss: 0.00040589438867755234
step: 150, loss: 0.00027910934295505285
step: 160, loss: 0.0025145397521555424
step: 170, loss: 0.04449588805437088
step: 180, loss: 0.0007441900670528412
step: 190, loss: 0.0018947263015434146
step: 200, loss: 0.0002858034858945757
step: 210, loss: 0.00022971928410697728
step: 220, loss: 0.0003667839046102017
step: 230, loss: 0.0036369659937918186
/usr/local/lib/python3.7/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=0.9853768278965129, f1=0.9887640449438202,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006575207808054984
step: 10, loss: 0.0019424293423071504
step: 20, loss: 0.000390388275263831
step: 30, loss: 0.00022162673121783882
step: 40, loss: 0.0003321169060654938
step: 50, loss: 0.001718503306619823
step: 60, loss: 0.014952156692743301
step: 70, loss: 0.07181493192911148
step: 80, loss: 0.0009230078430846334
step: 90, loss: 0.0004079567443113774
step: 100, loss: 0.000952509290073067
step: 110, loss: 0.08770088851451874
step: 120, loss: 0.008106586523354053
step: 130, loss: 0.013759749941527843
step: 140, loss: 0.007722408976405859
step: 150, loss: 0.00044702511513605714
step: 160, loss: 0.0008135873358696699
step: 170, loss: 0.0004557303327601403
step: 180, loss: 0.0006385442102327943
```

```
step: 190, loss: 0.001072289189323783
step: 200, loss: 0.001338408561423421
step: 210, loss: 0.0009355273214168847
step: 220, loss: 0.0016326821641996503
step: 230, loss: 0.00818548258394003
/usr/local/lib/python3.7/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=0.9854423292273236, f1=0.9821029082774049,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13496409356594086
step: 10, loss: 0.00021236446627881378
step: 20, loss: 0.0005560734425671399
step: 30, loss: 0.0010703395819291472
step: 40, loss: 0.0045325192622840405
step: 50, loss: 0.00020541147387120873
step: 60, loss: 0.002658756682649255
step: 70, loss: 0.0013252883218228817
step: 80, loss: 0.00021061699953861535
step: 90, loss: 0.002257630927488208
step: 100, loss: 0.07139492779970169
step: 110, loss: 0.007375766988843679
step: 120, loss: 0.002006979426369071
step: 130, loss: 0.0003994563012383878
step: 140, loss: 0.00067921652225777751
step: 150, loss: 0.0029192750807851553
step: 160, loss: 0.0006756323273293674
step: 170, loss: 0.0015479513676837087
step: 180, loss: 0.004196696449071169
step: 190, loss: 0.0017356837633997202
step: 200, loss: 0.014685221016407013
step: 210, loss: 0.00408443296328187
step: 220, loss: 0.00027031011995859444
step: 230, loss: 0.04959755763411522
/usr/local/lib/python3.7/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=0.9887640449438202, f1=0.9821826280623607,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.061533499509096146
step: 10, loss: 0.03804659843444824
step: 20, loss: 0.0014553772052749991
```

```
step: 30, loss: 0.00020093856437597424
step: 40, loss: 0.0006371235940605402
step: 50, loss: 0.002787250094115734
step: 60, loss: 0.004916273057460785
step: 70, loss: 0.00046719639794901013
step: 80, loss: 0.0028832186944782734
step: 90, loss: 0.0029365201480686665
step: 100, loss: 0.003506270470097661
step: 110, loss: 0.0005615268601104617
step: 120, loss: 9.152579877991229e-05
step: 130, loss: 0.0007378782029263675
step: 140, loss: 0.00040882843313738704
step: 150, loss: 0.0013157292269170284
step: 160, loss: 0.00014173737145029008
step: 170, loss: 0.0008680449682287872
step: 180, loss: 0.0023365276865661144
step: 190, loss: 6.945657514734194e-05
step: 200, loss: 0.001570162014104426
step: 210, loss: 0.0003536207368597388
step: 220, loss: 0.0003781883860938251
step: 230, loss: 0.006197354290634394
/usr/local/lib/python3.7/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=0.9887640449438202, f1=0.9741863075196409,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002742552289696157
step: 10, loss: 0.0009074048721233152
step: 20, loss: 0.0003481458406895399
step: 30, loss: 0.0009062237804755569
step: 40, loss: 7.74889558670111e-05
step: 50, loss: 9.324863640358672e-05
step: 60, loss: 0.0002408782602287829
step: 70, loss: 0.0014919084496796131
step: 80, loss: 0.048425547778606415
step: 90, loss: 0.017069226130843163
step: 100, loss: 0.002431433880701661
step: 110, loss: 0.0004608756280504167
step: 120, loss: 0.050509143620729446
step: 130, loss: 0.0002532981161493808
step: 140, loss: 0.0051688444800674915
step: 150, loss: 0.00013022596249356866
step: 160, loss: 0.001277240226045251
step: 170, loss: 0.0003268629079684615
step: 180, loss: 0.0005419357330538332
```

```
step: 190, loss: 0.0010936213657259941
step: 200, loss: 0.0008132155635394156
step: 210, loss: 0.00017094840586651117
step: 220, loss: 0.03926771879196167
step: 230, loss: 0.0009425032767467201
/usr/local/lib/python3.7/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=0.9876265466816648, f1=0.9776785714285714,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001757128193276003
step: 10, loss: 0.00020403462985996157
step: 20, loss: 0.0008199744042940438
step: 30, loss: 0.0018304046243429184
step: 40, loss: 0.00010589318844722584
step: 50, loss: 0.029618006199598312
step: 60, loss: 0.03620119020342827
step: 70, loss: 0.0009739737142808735
step: 80, loss: 0.00021632341668009758
step: 90, loss: 0.0008603634196333587
step: 100, loss: 0.000270493037533015
step: 110, loss: 0.0011055654613301158
step: 120, loss: 0.011377613991498947
step: 130, loss: 0.00034791018697433174
step: 140, loss: 0.03336218744516373
step: 150, loss: 7.128807919798419e-05
step: 160, loss: 0.0005935130175203085
step: 170, loss: 0.0006689022993668914
step: 180, loss: 0.0009024497121572495
step: 190, loss: 0.0010964112589135766
step: 200, loss: 8.105481538223103e-05
step: 210, loss: 0.000269620242761448
step: 220, loss: 0.00019995300681330264
step: 230, loss: 0.0005328114493750036
/usr/local/lib/python3.7/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=0.9876543209876544, f1=0.9820627802690582,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022349976643454283
step: 10, loss: 0.0001532344176666811
step: 20, loss: 8.151087968144566e-05
```

```
step: 30, loss: 0.00028078575269319117
step: 40, loss: 0.010558078065514565
step: 50, loss: 0.0048705292865633965
step: 60, loss: 0.00018844295118469745
step: 70, loss: 0.012864423915743828
step: 80, loss: 0.000672718626447022
step: 90, loss: 0.00031759956618770957
step: 100, loss: 0.0001928384881466627
step: 110, loss: 8.073062781477347e-05
step: 120, loss: 0.004250348079949617
step: 130, loss: 6.621894135605544e-05
step: 140, loss: 0.00046954338904470205
step: 150, loss: 6.666364788543433e-05
step: 160, loss: 0.00016060170310083777
step: 170, loss: 0.012307918630540371
step: 180, loss: 0.00048561551375314593
step: 190, loss: 0.015583917498588562
step: 200, loss: 8.002237882465124e-05
step: 210, loss: 0.0001301143056480214
step: 220, loss: 5.714888538932428e-05
step: 230, loss: 0.04420311003923416
/usr/local/lib/python3.7/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=0.9887640449438202, f1=0.977728285077951,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.594068640377373e-05
step: 10, loss: 0.00024840893456712365
step: 20, loss: 0.0001015674279187806
step: 30, loss: 0.00011229348456254229
step: 40, loss: 4.9962123739533126e-05
step: 50, loss: 6.805537850596011e-05
step: 60, loss: 0.006169418804347515
step: 70, loss: 9.889857756206766e-05
step: 80, loss: 5.850642628502101e-05
step: 90, loss: 4.064110908075236e-05
step: 100, loss: 7.212025229819119e-05
step: 110, loss: 5.2038634748896584e-05
step: 120, loss: 0.00019158587383572012
step: 130, loss: 0.0025867721997201443
step: 140, loss: 3.6659359466284513e-05
step: 150, loss: 0.0005287169478833675
step: 160, loss: 0.027581818401813507
step: 170, loss: 0.000187754281796515
step: 180, loss: 7.435183943016455e-05
```

```
step: 190, loss: 0.011743022128939629
step: 200, loss: 0.0036796436179429293
step: 210, loss: 5.193426113692112e-05
step: 220, loss: 0.012939782813191414
step: 230, loss: 6.184262747410685e-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=0.9865168539325843, f1=0.9798657718120806,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.437988824676722e-05
step: 10, loss: 7.897145405877382e-05
step: 20, loss: 7.625627040397376e-05
step: 30, loss: 0.019050512462854385
step: 40, loss: 0.00036389316665008664
step: 50, loss: 9.317112562712282e-05
step: 60, loss: 4.293184247217141e-05
step: 70, loss: 5.078083268017508e-05
step: 80, loss: 0.00032247137278318405
step: 90, loss: 4.3158637708984315e-05
step: 100, loss: 7.327867206186056e-05
step: 110, loss: 0.0004765275225508958
step: 120, loss: 0.0017611051443964243
step: 130, loss: 7.997608190635219e-05
step: 140, loss: 0.000640444690361619
step: 150, loss: 0.02503562904894352
step: 160, loss: 0.0002440354146528989
step: 170, loss: 8.437757060164586e-05
step: 180, loss: 0.0006038931314833462
step: 190, loss: 7.283477316377684e-05
step: 200, loss: 6.125299114501104e-05
step: 210, loss: 9.479219443164766e-05
step: 220, loss: 0.0003021334996446967
step: 230, loss: 0.00011165456817252561
/usr/local/lib/python3.7/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=0.9886877828054299, f1=0.9807474518686297,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.979588943067938e-05
step: 10, loss: 2.759257040452212e-05
step: 20, loss: 0.008928516879677773
```

```
step: 30, loss: 0.0003452568780630827
step: 40, loss: 4.5691787818213925e-05
step: 50, loss: 0.00018266677216161042
step: 60, loss: 3.548936729202978e-05
step: 70, loss: 4.779261871590279e-05
step: 80, loss: 6.25636603217572e-05
step: 90, loss: 0.00016159559891093522
step: 100, loss: 0.005307740066200495
step: 110, loss: 7.420059409923851e-05
step: 120, loss: 3.786973320529796e-05
step: 130, loss: 0.00010822355397976935
step: 140, loss: 0.00022798372083343565
step: 150, loss: 4.95140629936941e-05
step: 160, loss: 4.6725719585083425e-05
step: 170, loss: 5.354069435270503e-05
step: 180, loss: 4.842626367462799e-05
step: 190, loss: 0.0006548144156113267
step: 200, loss: 3.688267315737903e-05
step: 210, loss: 0.009160113520920277
step: 220, loss: 7.124522380763665e-05
step: 230, loss: 2.9648446798091754e-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=0.987598647125141, f1=0.9831649831649831,
best_f1=0.9854423292273236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.460632069618441e-05
step: 10, loss: 5.626598067465238e-05
step: 20, loss: 3.785502121900208e-05
step: 30, loss: 0.0030604610219597816
step: 40, loss: 4.8385263653472066e-05
step: 50, loss: 0.00037255254574120045
step: 60, loss: 0.0011997766559943557
step: 70, loss: 5.252122355159372e-05
step: 80, loss: 0.0013166648568585515
step: 90, loss: 2.602051245048642e-05
step: 100, loss: 4.2712355934781954e-05
step: 110, loss: 5.622773096547462e-05
step: 120, loss: 9.352466440759599e-05
step: 130, loss: 0.000466183468233794
step: 140, loss: 0.00013078034680802375
step: 150, loss: 0.00011916202493011951
step: 160, loss: 4.3366777390474454e-05
step: 170, loss: 5.816869452246465e-05
step: 180, loss: 5.004846752854064e-05
```

```
step: 190, loss: 0.008171826601028442
step: 200, loss: 3.218922938685864e-05
step: 210, loss: 0.03103950433433056
step: 220, loss: 0.0006140352925285697
step: 230, loss: 0.005530144087970257
/usr/local/lib/python3.7/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=0.9886877828054299, f1=0.9841986455981941,
best_f1=0.9854423292273236
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 318.85it/s]
load_f1 = 0.9864864864864865
real_f1 = 0.9853438556933484
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, 351.91it/s]
```

4.1.3 DBLP-GoogleScholar - Running the matcher

```python
# 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 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: 0.8005263805389404
step: 10, loss: 0.41542503237724304
step: 20, loss: 0.4751308858394623
step: 30, loss: 0.35599035024642944
step: 40, loss: 0.2615565061569214
step: 50, loss: 0.1311662346124649
step: 60, loss: 0.09639929980039597
step: 70, loss: 0.128691628575325
step: 80, loss: 0.02877247892320156
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.10919712483882904
step: 100, loss: 0.3552566170692444
step: 110, loss: 0.05012024939060211
step: 120, loss: 0.03894366696476936
step: 130, loss: 0.03445380926132202
step: 140, loss: 0.20342908799648285
step: 150, loss: 0.05708911642432213
step: 160, loss: 0.1577497273683548
step: 170, loss: 0.15568400919437408
step: 180, loss: 0.09928257763385773
step: 190, loss: 0.13670258224010468
step: 200, loss: 0.09894596040248871
step: 210, loss: 0.08523791283369064
```

```
step: 220, loss: 0.08513809740543365
step: 230, loss: 0.06626058369874954
step: 240, loss: 0.04156443476676941
step: 250, loss: 0.04273556172847748
step: 260, loss: 0.020873621106147766
step: 270, loss: 0.02002468705177307
step: 280, loss: 0.1253444254398346
step: 290, loss: 0.05394814535975456
step: 300, loss: 0.17237399518489838
step: 310, loss: 0.024959636852145195
step: 320, loss: 0.19484050571918488
step: 330, loss: 0.09039602428674698
step: 340, loss: 0.16943830251693726
step: 350, loss: 0.0738149955868721
step: 360, loss: 0.09765815734863281
step: 370, loss: 0.13475622236728668
step: 380, loss: 0.19343321025371552
step: 390, loss: 0.05437782406806946
step: 400, loss: 0.03819500282406807
step: 410, loss: 0.010391340591013432
step: 420, loss: 0.007714881096035242
step: 430, loss: 0.04238872975111008
step: 440, loss: 0.05146821588277817
step: 450, loss: 0.0212898887693882
step: 460, loss: 0.1617482751607895
step: 470, loss: 0.20933009684085846
step: 480, loss: 0.2532021105289459
step: 490, loss: 0.0217225793749094
step: 500, loss: 0.009493556804955006
step: 510, loss: 0.10463038831949234
step: 520, loss: 0.006124683655798435
step: 530, loss: 0.07914814352989197
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9374416433239963, f1=0.9335205992509362,
best_f1=0.9335205992509362
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08435879647731781
step: 10, loss: 0.155759796500206
step: 20, loss: 0.14314350485801697
step: 30, loss: 0.050835005939006805
```

```
step: 40, loss: 0.0038190968334674835
step: 50, loss: 0.09805994480848312
step: 60, loss: 0.06500545144081116
step: 70, loss: 0.13375574350357056
step: 80, loss: 0.008593018166720867
step: 90, loss: 0.0072690569795668125
step: 100, loss: 0.34574517607688904
step: 110, loss: 0.028621042147278786
step: 120, loss: 0.05738678574562073
step: 130, loss: 0.03752998635172844
step: 140, loss: 0.023567043244838715
step: 150, loss: 0.09234211593866348
step: 160, loss: 0.03746102750301361
step: 170, loss: 0.14624091982841492
step: 180, loss: 0.0018869559280574322
step: 190, loss: 0.02362917922437191
step: 200, loss: 0.02520080655813217
step: 210, loss: 0.03895355761051178
step: 220, loss: 0.18731562793254852
step: 230, loss: 0.04077160730957985
step: 240, loss: 0.11369020491838455
step: 250, loss: 0.04439351707696915
step: 260, loss: 0.005501703359186649
step: 270, loss: 0.16104674339294434
step: 280, loss: 0.20342324674129486
step: 290, loss: 0.1348351389169693
step: 300, loss: 0.1438322365283966
step: 310, loss: 0.14200620353221893
step: 320, loss: 0.1325550526380539
step: 330, loss: 0.07723671942949295
step: 340, loss: 0.020009897649288177
step: 350, loss: 0.053088001906871796
step: 360, loss: 0.008371756412088871
step: 370, loss: 0.004730198998004198
step: 380, loss: 0.1356775164604187
step: 390, loss: 0.025342712178826332
step: 400, loss: 0.07031790912151337
step: 410, loss: 0.0024383857380598783
step: 420, loss: 0.02009551040828228
step: 430, loss: 0.014934390783309937
step: 440, loss: 0.01190318534741402
step: 450, loss: 0.02033355087041855
step: 460, loss: 0.1391797959804535
step: 470, loss: 0.039629414677619934
step: 480, loss: 0.4096832275390625
step: 490, loss: 0.04237288236618042
step: 500, loss: 0.011495371349155903
step: 510, loss: 0.03075646422803402
```

```
step: 520, loss: 0.08122538030147552
step: 530, loss: 0.07389518618583679
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9413394919168592, f1=0.9411219286045434,
best_f1=0.9411219286045434
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04023981839418411
step: 10, loss: 0.036933835595846176
step: 20, loss: 0.12592971324920654
step: 30, loss: 0.1710810363292694
step: 40, loss: 0.005721107590943575
step: 50, loss: 0.019242750480771065
step: 60, loss: 0.0032702255994081497
step: 70, loss: 0.05656324326992035
step: 80, loss: 0.014685891568660736
step: 90, loss: 0.07026880234479904
step: 100, loss: 0.010431163012981415
step: 110, loss: 0.004883939400315285
step: 120, loss: 0.2793063819408417
step: 130, loss: 0.07737153023481369
step: 140, loss: 0.04097018763422966
step: 150, loss: 0.04028654471039772
step: 160, loss: 0.004974490962922573
step: 170, loss: 0.0060089873149991035
step: 180, loss: 0.07320134341716766
step: 190, loss: 0.07509499788284302
step: 200, loss: 0.022261101752519608
step: 210, loss: 0.02351999841630459
step: 220, loss: 0.02653687819838524
step: 230, loss: 0.07159139215946198
step: 240, loss: 0.01541369128972292
step: 250, loss: 0.009660662151873112
step: 260, loss: 0.03161406144499779
step: 270, loss: 0.0023001518566161394
step: 280, loss: 0.0033570146188139915
step: 290, loss: 0.026544049382209778
step: 300, loss: 0.06346810609102249
step: 310, loss: 0.07187255471944809
step: 320, loss: 0.13707077503204346
step: 330, loss: 0.002037402242422104
```

```
step: 340, loss: 0.0011326177045702934
step: 350, loss: 0.00913529098033905
step: 360, loss: 0.006673912983387709
step: 370, loss: 0.0010361397871747613
step: 380, loss: 0.0036876555532217026
step: 390, loss: 0.02699778601527214
step: 400, loss: 0.06501171737909317
step: 410, loss: 0.006718612276017666
step: 420, loss: 0.04491594061255455
step: 430, loss: 0.19572550058364868
step: 440, loss: 0.04136955738067627
step: 450, loss: 0.05967304855585098
step: 460, loss: 0.16638848185539246
step: 470, loss: 0.004523608833551407
step: 480, loss: 0.0036352286115288734
step: 490, loss: 0.006607538089156151
step: 500, loss: 0.11451689153909683
step: 510, loss: 0.0069505600258708
step: 520, loss: 0.031013237312436104
step: 530, loss: 0.007839917205274105
/usr/local/lib/python3.7/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=0.9377817853922452, f1=0.9394347240915207,
best_f1=0.9411219286045434
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002620538929477334
step: 10, loss: 0.08371343463659286
step: 20, loss: 0.038151636719703674
step: 30, loss: 0.0032545635476708412
step: 40, loss: 0.007674440275877714
step: 50, loss: 0.003252844326198101
step: 60, loss: 0.001427200622856617
step: 70, loss: 0.002966254483908415
step: 80, loss: 0.09385128319263458
step: 90, loss: 0.060923825949430466
step: 100, loss: 0.12888574600219727
step: 110, loss: 0.04532938823103905
step: 120, loss: 0.009268616326153278
step: 130, loss: 0.041480179876089096
step: 140, loss: 0.02406330220401287
step: 150, loss: 0.029415572062134743
step: 160, loss: 0.002006766851991415
step: 170, loss: 0.0016437794547528028
step: 180, loss: 0.004602559842169285
step: 190, loss: 0.05172500014305115
```

```
step: 200, loss: 0.0007729352219030261
step: 210, loss: 0.06897074729204178
step: 220, loss: 0.0016083435621112585
step: 230, loss: 0.2358080893754959
step: 240, loss: 0.0026967176236212254
step: 250, loss: 0.021422095596790314
step: 260, loss: 0.09034053981304169
step: 270, loss: 0.013535469770431519
step: 280, loss: 0.002798102330416441
step: 290, loss: 0.026575999334454536
step: 300, loss: 0.021059434860944748
step: 310, loss: 0.016991496086120605
step: 320, loss: 0.07830539345741272
step: 330, loss: 0.08722801506519318
step: 340, loss: 0.03868700563907623
step: 350, loss: 0.002205859636887908
step: 360, loss: 0.015270276926457882
step: 370, loss: 0.004375205375254154
step: 380, loss: 0.0025413960684090853
step: 390, loss: 0.044029515236616135
step: 400, loss: 0.0050577521324157715
step: 410, loss: 0.043366167694330215
step: 420, loss: 0.007078452967107296
step: 430, loss: 0.01535546500235796
step: 440, loss: 0.16968446969985962
step: 450, loss: 0.05386173352599144
step: 460, loss: 0.013590057380497456
step: 470, loss: 0.009621571749448776
step: 480, loss: 0.0029254844412207603
step: 490, loss: 0.030092507600784302
step: 500, loss: 0.012970651499927044
step: 510, loss: 0.024317802861332893
step: 520, loss: 0.0027100383304059505
step: 530, loss: 0.001884161145426333
/usr/local/lib/python3.7/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=0.9380530973451328, f1=0.9410681399631675,
best_f1=0.9411219286045434
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005730379838496447
step: 10, loss: 0.001750613795593381
step: 20, loss: 0.0004841024347115308
step: 30, loss: 0.004255830310285091
step: 40, loss: 0.033136866986751556
step: 50, loss: 0.004472510423511267
```

```
step: 60, loss: 0.006230528932064772
step: 70, loss: 0.00283241830766201
step: 80, loss: 0.0008097944664768875
step: 90, loss: 0.017365064471960068
step: 100, loss: 0.19897378981113434
step: 110, loss: 0.005690527148544788
step: 120, loss: 0.00555471470579505
step: 130, loss: 0.002473689615726471
step: 140, loss: 0.004807351157069206
step: 150, loss: 0.003778624115511775
step: 160, loss: 0.003084537573158741
step: 170, loss: 0.020859450101852417
step: 180, loss: 0.0009937871946021914
step: 190, loss: 0.0008670722600072622
step: 200, loss: 0.01508663035929203
step: 210, loss: 0.0027692506555467844
step: 220, loss: 0.002840849803760478
step: 230, loss: 0.0017363756196573377
step: 240, loss: 0.040137503296136856
step: 250, loss: 0.00044430760317482054
step: 260, loss: 0.01626964844763279
step: 270, loss: 0.0880109891295433
step: 280, loss: 0.10632544755935669
step: 290, loss: 0.0017883627442643046
step: 300, loss: 0.022881144657731056
step: 310, loss: 0.005904076620936394
step: 320, loss: 0.0034702590201050043
step: 330, loss: 0.0006891351076774299
step: 340, loss: 0.002650150330737233
step: 350, loss: 0.013388545252382755
step: 360, loss: 0.0025419851299375296
step: 370, loss: 0.006596776191145182
step: 380, loss: 0.0430387444794178
step: 390, loss: 0.0036806240677833557
step: 400, loss: 0.15858900547027588
step: 410, loss: 0.0013247760944068432
step: 420, loss: 0.006831660401076078
step: 430, loss: 0.00581190874800086
step: 440, loss: 0.028964392840862274
step: 450, loss: 0.00922250747680664
step: 460, loss: 0.010997605510056019
step: 470, loss: 0.07240884751081467
step: 480, loss: 0.02536497823894024
step: 490, loss: 0.0012028667842969298
step: 500, loss: 0.0019247207092121243
step: 510, loss: 0.006243873853236437
step: 520, loss: 0.00146814517211169
step: 530, loss: 0.058792538940906525
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9416705552963137, f1=0.937988056959118,
best_f1=0.937988056959118
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018328482983633876
step: 10, loss: 0.024371366947889328
step: 20, loss: 0.0035388146061450243
step: 30, loss: 0.03858516737818718
step: 40, loss: 0.008784573525190353
step: 50, loss: 0.03674693778157234
step: 60, loss: 0.00347258266992867
step: 70, loss: 0.024699687957763672
step: 80, loss: 0.012926755473017693
step: 90, loss: 0.0009919159347191453
step: 100, loss: 0.0009907725034281611
step: 110, loss: 0.006631914060562849
step: 120, loss: 0.0009817179525271058
step: 130, loss: 0.0005821932572871447
step: 140, loss: 0.0007279480341821909
step: 150, loss: 0.011298674158751965
step: 160, loss: 0.0002103893057210371
step: 170, loss: 0.00021720165386795998
step: 180, loss: 0.00658212136477232
step: 190, loss: 0.017896654084324837
step: 200, loss: 0.00010494231537450105
step: 210, loss: 0.00018821119738277048
step: 220, loss: 0.003627397818490863
step: 230, loss: 0.0006896972190588713
step: 240, loss: 0.0017364986706525087
step: 250, loss: 0.009582470171153545
step: 260, loss: 0.0019816996064037085
step: 270, loss: 0.012614169158041477
step: 280, loss: 0.01961461827158928
step: 290, loss: 0.0018979123560711741
step: 300, loss: 0.014237459748983383
step: 310, loss: 0.007765790447592735
step: 320, loss: 0.057096172124147415
step: 330, loss: 0.0013530370779335499
step: 340, loss: 0.014894142746925354
step: 350, loss: 0.09487202763557434
```

```
step: 360, loss: 0.047719817608594894
step: 370, loss: 0.0003060655726585537
step: 380, loss: 0.016979817301034927
step: 390, loss: 0.005099428817629814
step: 400, loss: 0.0005940985283814371
step: 410, loss: 0.00012002299627056345
step: 420, loss: 0.16266292333602905
step: 430, loss: 0.0002513022336643189
step: 440, loss: 0.05157158523797989
step: 450, loss: 0.016904065385460854
step: 460, loss: 0.004742344841361046
step: 470, loss: 0.22928781807422638
step: 480, loss: 0.044144004583358765
step: 490, loss: 0.0022947778925299644
step: 500, loss: 0.0017709919484332204
step: 510, loss: 0.018418796360492706
step: 520, loss: 0.020038045942783356
step: 530, loss: 0.005559171084314585
/usr/local/lib/python3.7/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=0.9369708372530574, f1=0.936249418334109,
best_f1=0.937988056959118
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015681985765695572
step: 10, loss: 0.011838540434837341
step: 20, loss: 0.019974227994680405
step: 30, loss: 0.07290226221084595
step: 40, loss: 0.006009502802044153
step: 50, loss: 0.0013964232057332993
step: 60, loss: 0.0005199272418394685
step: 70, loss: 0.002168983453884721
step: 80, loss: 0.00039002191624604166
step: 90, loss: 0.0030108727514743805
step: 100, loss: 0.004989616572856903
step: 110, loss: 0.055993515998125076
step: 120, loss: 0.009247093461453915
step: 130, loss: 0.0014638114953413606
step: 140, loss: 0.017384523525834084
step: 150, loss: 0.0005131965735927224
step: 160, loss: 0.001430589472874999
step: 170, loss: 0.0002269282704219222
step: 180, loss: 0.0006833773804828525
step: 190, loss: 0.02326984517276287
step: 200, loss: 0.03840944916009903
step: 210, loss: 0.0017541844863444567
```

```
step: 220, loss: 0.0013894860167056322
step: 230, loss: 0.001300309901125729
step: 240, loss: 0.011250438168644905
step: 250, loss: 0.060012564063072205
step: 260, loss: 0.018193598836660385
step: 270, loss: 0.004588533658534288
step: 280, loss: 0.01214238628745079
step: 290, loss: 0.004409424029290676
step: 300, loss: 0.0008682553307153285
step: 310, loss: 0.0007152078906074166
step: 320, loss: 0.030207309871912003
step: 330, loss: 0.003137449501082301
step: 340, loss: 0.0336049385368824
step: 350, loss: 0.000536877429112792
step: 360, loss: 0.003721467684954405
step: 370, loss: 0.0005720655317418277
step: 380, loss: 0.0008547900361008942
step: 390, loss: 0.0010777339339256287
step: 400, loss: 0.0014900292735546827
step: 410, loss: 0.002007852541282773
step: 420, loss: 0.00027051838696934283
step: 430, loss: 0.00020643090829253197
step: 440, loss: 0.00014550394553225487
step: 450, loss: 0.00010423745698062703
step: 460, loss: 0.012450363487005234
step: 470, loss: 0.1080687940120697
step: 480, loss: 0.012303823605179787
step: 490, loss: 0.000617774436250329
step: 500, loss: 0.00038666650652885437
step: 510, loss: 0.016542768105864525
step: 520, loss: 0.0024952655658125877
step: 530, loss: 0.00016331484948750585
/usr/local/lib/python3.7/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=0.9394919168591224, f1=0.9372693726937269,
best_f1=0.937988056959118
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020494451746344566
step: 10, loss: 0.04779642075300217
step: 20, loss: 0.011133551597595215
step: 30, loss: 0.018713993951678276
step: 40, loss: 0.0003119935863651335
step: 50, loss: 0.0002578071434982121
step: 60, loss: 0.01623334176838398
step: 70, loss: 0.006062277127057314
```

```
step: 80, loss: 0.003415636718273163
step: 90, loss: 0.00021673805895261467
step: 100, loss: 0.0034384122118353844
step: 110, loss: 0.0008068797178566456
step: 120, loss: 0.14895085990428925
step: 130, loss: 0.030824018642306328
step: 140, loss: 6.410917558241636e-05
step: 150, loss: 0.0010971801821142435
step: 160, loss: 0.0023809128906577826
step: 170, loss: 0.0002942870487459004
step: 180, loss: 0.0010003920178860426
step: 190, loss: 0.0006047164788469672
step: 200, loss: 0.004950840026140213
step: 210, loss: 0.003920793998986483
step: 220, loss: 0.10067486017942429
step: 230, loss: 0.002577271545305848
step: 240, loss: 0.0012839942937716842
step: 250, loss: 0.009732391685247421
step: 260, loss: 0.014054571278393269
step: 270, loss: 8.220221207011491e-05
step: 280, loss: 0.00021167185332160443
step: 290, loss: 0.0004089061403647065
step: 300, loss: 0.00034655415220186114
step: 310, loss: 0.02944369800388813
step: 320, loss: 0.00047835250734351575
step: 330, loss: 0.021074384450912476
step: 340, loss: 0.0015486768679693341
step: 350, loss: 0.13039326667785645
step: 360, loss: 0.000618667050730437
step: 370, loss: 0.0011733826249837875
step: 380, loss: 0.008166909217834473
step: 390, loss: 0.0005295485025271773
step: 400, loss: 0.07211387902498245
step: 410, loss: 0.0006784027791582048
step: 420, loss: 0.0016600240487605333
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 430, loss: 0.0005647815414704382
step: 440, loss: 0.0013529807329177856
step: 450, loss: 0.002784262876957655
step: 460, loss: 0.0018664461094886065
step: 470, loss: 0.001231941394507885
step: 480, loss: 0.0036343391984701157
step: 490, loss: 0.0004595079808495939
step: 500, loss: 0.0015437209513038397
step: 510, loss: 0.00038510659942403436
step: 520, loss: 0.002281543333083391
step: 530, loss: 0.0003952353144995868
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9444444444444445, f1=0.9429864253393664,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010380590101704001
step: 10, loss: 0.0005307439132593572
step: 20, loss: 0.000799145083874464
step: 30, loss: 0.0004491645086091012
step: 40, loss: 0.001050864695571363
step: 50, loss: 0.0019971479196101427
step: 60, loss: 7.803470361977816e-05
step: 70, loss: 0.14220066368579865
step: 80, loss: 0.00023836587206460536
step: 90, loss: 0.0006961271283216774
step: 100, loss: 9.383080760017037e-05
step: 110, loss: 0.0005176051636226475
step: 120, loss: 0.04557617008686066
step: 130, loss: 0.003536288160830736
step: 140, loss: 0.001734166988171637
step: 150, loss: 0.000147067810758017
step: 160, loss: 0.0009190560085698962
step: 170, loss: 0.000490392732899636
step: 180, loss: 0.00033712008735165
step: 190, loss: 0.148972749710083
step: 200, loss: 0.00898121390491724
step: 210, loss: 0.001224365783855319
step: 220, loss: 0.0009875008836388588
step: 230, loss: 0.0003163847723044455
step: 240, loss: 0.0014917217195034027
step: 250, loss: 0.013601726852357388
step: 260, loss: 0.0010151732712984085
step: 270, loss: 0.0033650831319391727
step: 280, loss: 0.001400059089064598
step: 290, loss: 0.000303196138702333
step: 300, loss: 0.003514179727062583
step: 310, loss: 9.147108357865363e-05
step: 320, loss: 0.00043325888691470027
step: 330, loss: 0.0009196047903969884
step: 340, loss: 0.0007796260761097074
step: 350, loss: 0.0002329065027879551
step: 360, loss: 0.025524422526359558

```
step: 370, loss: 0.0005796998739242554
step: 380, loss: 0.0006330094765871763
step: 390, loss: 0.00847314391285181
step: 400, loss: 0.0005678705056197941
step: 410, loss: 0.0002188187645515427
step: 420, loss: 0.0033944030292332172
step: 430, loss: 5.72755016037263e-05
step: 440, loss: 0.00028906058287248015
step: 450, loss: 0.00016324482567142695
step: 460, loss: 0.0008941454580053687
step: 470, loss: 0.0001542923564556986
step: 480, loss: 0.0037426615599542856
step: 490, loss: 0.024057749658823013
step: 500, loss: 0.004053478129208088
step: 510, loss: 0.0261154193431139
step: 520, loss: 0.0005032025510445237
step: 530, loss: 0.0015002924483269453
/usr/local/lib/python3.7/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=0.9389846297158826, f1=0.9383057090239412,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007000192999839783
step: 10, loss: 0.0010369764640927315
step: 20, loss: 0.00061649305280298
step: 30, loss: 0.0009441343718208373
step: 40, loss: 0.0010148031869903207
step: 50, loss: 8.033713675104082e-05
step: 60, loss: 0.00013513886369764805
step: 70, loss: 0.0026219661813229322
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.0011138509726151824
step: 90, loss: 0.013558406382799149
step: 100, loss: 0.0012217869516462088
step: 110, loss: 0.002315846271812916
step: 120, loss: 0.14274375140666962
step: 130, loss: 0.00014754677249584347
step: 140, loss: 0.022488616406917572
step: 150, loss: 0.012366737239062786
step: 160, loss: 0.0009605035302229226
step: 170, loss: 0.00100520602427423
step: 180, loss: 0.005688668228685856
step: 190, loss: 0.0023265888448804617
step: 200, loss: 0.0003717833897098899
step: 210, loss: 0.001135646947659552
```

```
step: 220, loss: 0.008931319229320479
step: 230, loss: 0.0009864472085610032
step: 240, loss: 7.5006035331171e-05
step: 250, loss: 0.001584170968271792
step: 260, loss: 0.010684346780180931
step: 270, loss: 0.0014813541201874614
step: 280, loss: 8.32822552183643e-05
step: 290, loss: 0.0001809145323932171
step: 300, loss: 0.021485570818185806
step: 310, loss: 0.00069830642314622777
step: 320, loss: 0.000828960386570543
step: 330, loss: 0.002510746708139777
step: 340, loss: 0.0011985779274255037
step: 350, loss: 0.0005290994886308908
step: 360, loss: 0.00033646743395365775
step: 370, loss: 0.0011964502045884728
step: 380, loss: 0.000304180575767532
step: 390, loss: 0.00027512534870766103
step: 400, loss: 0.00011592281225603074
step: 410, loss: 0.0016814367845654488
step: 420, loss: 0.045721184462308884
step: 430, loss: 0.0003497414290904999
step: 440, loss: 9.228064300259575e-05
step: 450, loss: 0.005246104672551155
step: 460, loss: 0.02137441374361515
step: 470, loss: 4.181471376796253e-05
step: 480, loss: 8.185082697309554e-05
step: 490, loss: 0.14299483597278595
step: 500, loss: 0.015202910639345646
step: 510, loss: 0.0007133527542464435
step: 520, loss: 0.00038603669963777065
step: 530, loss: 0.017932947725057602
/usr/local/lib/python3.7/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=0.9423165666820489, f1=0.9431192660550459,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.036365363746881485
step: 10, loss: 0.008114946074783802
step: 20, loss: 0.00019234305364079773
step: 30, loss: 0.0025282949209213257
step: 40, loss: 0.002972945338115096
step: 50, loss: 0.006812617182731628
step: 60, loss: 0.00011406794510548934
step: 70, loss: 4.091527807759121e-05
```

```
step: 80, loss: 0.00041602333658374846
step: 90, loss: 0.005095477681607008
step: 100, loss: 0.0003122941998299211
step: 110, loss: 0.00013298357953317463
step: 120, loss: 0.004714332986623049
step: 130, loss: 0.0002141363947885111
step: 140, loss: 0.00016980456712190062
step: 150, loss: 0.0011356439208611846
step: 160, loss: 0.0010825522476807237
step: 170, loss: 0.06458991765975952
step: 180, loss: 0.0006352917407639325
step: 190, loss: 0.0009093563421629369
step: 200, loss: 0.003440307918936014
step: 210, loss: 0.0002904111461248249
step: 220, loss: 0.001260519027709961
step: 230, loss: 0.0012138350866734982
step: 240, loss: 3.84844170184806e-05
step: 250, loss: 0.00014192977687343955
step: 260, loss: 0.00016632092592772096
step: 270, loss: 0.002253302140161395
step: 280, loss: 0.00017397770716343075
step: 290, loss: 0.013994205743074417
step: 300, loss: 0.01376517303287983
step: 310, loss: 0.001027330057695508
step: 320, loss: 0.017051726579666138
step: 330, loss: 0.0018980078166350722
step: 340, loss: 0.0005255891592241824
step: 350, loss: 0.00012303584662731737
step: 360, loss: 4.366526991361752e-05
step: 370, loss: 4.6766785089857876e-05
step: 380, loss: 8.488872845191509e-05
step: 390, loss: 0.0089088324457407
step: 400, loss: 8.903419802663848e-05
step: 410, loss: 0.017705494537949562
step: 420, loss: 0.0009460880537517369
step: 430, loss: 0.00016019221220631152
step: 440, loss: 0.0003042830794584006
step: 450, loss: 9.888863132800907e-05
step: 460, loss: 0.0012985795037820935
step: 470, loss: 0.0005965505843050778
step: 480, loss: 0.00173462787643075
step: 490, loss: 0.03763331100344658
step: 500, loss: 0.00013432645937427878
step: 510, loss: 9.743680857354775e-05
step: 520, loss: 4.9748206947697327e-05
step: 530, loss: 8.241987234214321e-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=0.9364928909952607, f1=0.9320754716981132,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005472052725963295
step: 10, loss: 0.0008773346780799329
step: 20, loss: 4.4440497731557116e-05
step: 30, loss: 0.00457368977367878
step: 40, loss: 0.00013312272494658828
step: 50, loss: 0.0016156930942088366
step: 60, loss: 0.0002394976036157459
step: 70, loss: 0.17887575924396515
step: 80, loss: 0.002485139761120081
step: 90, loss: 0.016239961609244347
step: 100, loss: 0.002093004994094372
step: 110, loss: 0.00014656345592811704
step: 120, loss: 0.0003032584208995104
step: 130, loss: 0.00010522667434997857
step: 140, loss: 0.0013511335710063577
step: 150, loss: 0.004634852521121502
step: 160, loss: 0.0010637465165928006
step: 170, loss: 0.0001403950445819646
step: 180, loss: 0.00020636132103390992
step: 190, loss: 9.515699639450759e-05
step: 200, loss: 6.742133700754493e-05
step: 210, loss: 0.0007111418526619673
step: 220, loss: 0.0008379664504900575
step: 230, loss: 0.001842255238443613
step: 240, loss: 0.00025926914531737566
step: 250, loss: 8.858376531861722e-05
step: 260, loss: 0.0018880442949011922
step: 270, loss: 0.0018092578975483775
step: 280, loss: 5.388560748542659e-05
step: 290, loss: 0.005078032147139311
step: 300, loss: 0.13126572966575623
step: 310, loss: 0.00040070447721518576
step: 320, loss: 0.01100926660001278
step: 330, loss: 0.0009429882629774511
step: 340, loss: 0.013684109784662724
step: 350, loss: 0.016479015350341797
step: 360, loss: 0.0012239165371283889
step: 370, loss: 8.034630445763469e-05
step: 380, loss: 0.00019689499458763748
step: 390, loss: 5.0991930038435385e-05
step: 400, loss: 0.0004330045194365084
step: 410, loss: 0.00809506792575121
```

```
step: 420, loss: 0.0007507638074457645
step: 430, loss: 0.000285808666376397
step: 440, loss: 0.0009369925828650594
step: 450, loss: 0.0012217387557029724
step: 460, loss: 0.012395313940942287
step: 470, loss: 6.9548434112221e-05
step: 480, loss: 3.617007314460352e-05
step: 490, loss: 0.00017592089716345072
step: 500, loss: 0.13701346516609192
step: 510, loss: 0.00020516289805527776
step: 520, loss: 0.01769889146089554
step: 530, loss: 0.00023809012782294303
/usr/local/lib/python3.7/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=0.9375293565054016, f1=0.9386416861826699,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012327057775110006
step: 10, loss: 0.00012865461758337915
step: 20, loss: 0.024728043004870415
step: 30, loss: 0.0003167492395732552
step: 40, loss: 0.00022169403382577002
step: 50, loss: 6.179466436151415e-05
step: 60, loss: 0.00033625701325945556
step: 70, loss: 0.00018401604029349983
step: 80, loss: 0.0001006919119390659
step: 90, loss: 0.0010578096844255924
step: 100, loss: 0.00014767130778636783
step: 110, loss: 0.03734869509935379
step: 120, loss: 0.00019205921853426844
step: 130, loss: 0.0018094491679221392
step: 140, loss: 0.06286241114139557
step: 150, loss: 0.0009951300453394651
step: 160, loss: 0.0026368708349764347
step: 170, loss: 0.0049406434409320354
step: 180, loss: 0.000275130762020126
step: 190, loss: 7.891570567153394e-05
step: 200, loss: 0.0056563979014754295
step: 210, loss: 0.0022858998272567987
step: 220, loss: 0.00013417084119282663
step: 230, loss: 0.00020038927323184907
step: 240, loss: 2.7364481866243295e-05
step: 250, loss: 0.04357704892754555
step: 260, loss: 0.0050620888359844685
step: 270, loss: 0.00018057046690955758
```

```
step: 280, loss: 2.7819338356493972e-05
step: 290, loss: 0.1182367280125618
step: 300, loss: 0.00034718672395683825
step: 310, loss: 0.0032139441464096308
step: 320, loss: 3.792801362578757e-05
step: 330, loss: 0.003467452945187688
step: 340, loss: 0.00016621890245005488
step: 350, loss: 0.03170472010970116
step: 360, loss: 9.844728629104793e-05
step: 370, loss: 7.415620348183438e-05
step: 380, loss: 0.00044200941920280457
step: 390, loss: 0.00014728907262906432
step: 400, loss: 6.0956594097660854e-05
step: 410, loss: 0.0007898050826042891
step: 420, loss: 0.00028724598814733326
step: 430, loss: 0.0011361943325027823
step: 440, loss: 0.0035212580114603043
step: 450, loss: 0.00016119510110000201
step: 460, loss: 2.0823774320888333e-05
step: 470, loss: 0.019834032282233238
step: 480, loss: 0.001009408850222826
step: 490, loss: 8.91496310941875e-05
step: 500, loss: 0.0016351834638044238
step: 510, loss: 7.46251389500685e-05
step: 520, loss: 0.0014980166452005506
step: 530, loss: 0.00122172012925148
/usr/local/lib/python3.7/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=0.9387186629526463, f1=0.9400921658986175,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00019172686734236777
step: 10, loss: 8.591202640673146e-05
step: 20, loss: 1.917748522828333e-05
step: 30, loss: 0.00021140392345841974
step: 40, loss: 0.00069882080424577
step: 50, loss: 0.0002693166316021234
step: 60, loss: 3.4531210985733196e-05
step: 70, loss: 3.6739627830684185e-05
step: 80, loss: 0.00403297645971179
step: 90, loss: 2.311043135705404e-05
step: 100, loss: 0.00016277021495625377
step: 110, loss: 3.17372105200775e-05
step: 120, loss: 7.495946192648262e-05
step: 130, loss: 0.003600429743528366
```

```
step: 140, loss: 0.061722081154584885
step: 150, loss: 0.0009161225170828402
step: 160, loss: 0.001702010165899992
step: 170, loss: 0.0020223241299390793
step: 180, loss: 0.0014091074699535966
step: 190, loss: 9.907661296892911e-05
step: 200, loss: 0.0012781412806361914
step: 210, loss: 0.0004881888162344694
step: 220, loss: 0.000539381115231663
step: 230, loss: 0.0017353304428979754
step: 240, loss: 0.0005700589972548187
step: 250, loss: 5.3877331083640456e-05
step: 260, loss: 1.827601226978004e-05
step: 270, loss: 0.004932384006679058
step: 280, loss: 4.476710455492139e-05
step: 290, loss: 2.6217188860755414e-05
step: 300, loss: 7.170718890847638e-05
step: 310, loss: 0.014023211784660816
step: 320, loss: 8.513725333614275e-05
step: 330, loss: 0.031468648463487625
step: 340, loss: 6.567981472471729e-05
step: 350, loss: 0.017678391188383102
step: 360, loss: 0.002438341034576297
step: 370, loss: 0.00017364609811920673
step: 380, loss: 0.0005854410701431334
step: 390, loss: 0.0010687952162697911
step: 400, loss: 3.213874151697382e-05
step: 410, loss: 7.026113598840311e-05
step: 420, loss: 8.271675324067473e-05
step: 430, loss: 0.00022926554083824158
step: 440, loss: 0.00010785911581479013
step: 450, loss: 0.0002243541384814307
step: 460, loss: 0.0016558060888200998
step: 470, loss: 2.9874152460251935e-05
step: 480, loss: 1.852190871431958e-05
step: 490, loss: 0.0001870648848125711
step: 500, loss: 0.0005478411330841482
step: 510, loss: 3.378767360118218e-05
step: 520, loss: 2.1054909666418098e-05
step: 530, loss: 7.284684397745878e-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=0.9413394919168592, f1=0.9423076923076922,
best_f1=0.9429864253393664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.00016295895329676569
step: 10, loss: 0.0003225776308681816
step: 20, loss: 3.626128818723373e-05
step: 30, loss: 0.0006661959923803806
step: 40, loss: 0.0033829775638878345
step: 50, loss: 0.00046978454338386655
step: 60, loss: 2.924204818555154e-05
step: 70, loss: 0.00032570710754953325
step: 80, loss: 2.0280200260458514e-05
step: 90, loss: 2.082691753457766e-05
step: 100, loss: 1.5444884411408566e-05
step: 110, loss: 0.004363395273685455
step: 120, loss: 0.0022002840414643288
step: 130, loss: 0.0012986217625439167
step: 140, loss: 1.685672214080114e-05
step: 150, loss: 0.005583038553595543
step: 160, loss: 8.072037599049509e-05
step: 170, loss: 0.013499513268470764
step: 180, loss: 0.008079413324594498
step: 190, loss: 0.0005366649711504579
step: 200, loss: 0.0026048922445625067
step: 210, loss: 0.00013841292820870876
step: 220, loss: 3.4765369491651654e-05
step: 230, loss: 0.001192459836602211
step: 240, loss: 0.0029336444567888975
step: 250, loss: 0.00039189966628327966
step: 260, loss: 0.00010172020847676322
step: 270, loss: 0.0011680652387440205
step: 280, loss: 0.0016612386098131537
step: 290, loss: 0.00332314008846879
step: 300, loss: 0.03351437300443649
step: 310, loss: 0.002427659695968032
step: 320, loss: 5.721509660361335e-05
step: 330, loss: 0.00029269952210597694
step: 340, loss: 0.00037962099304422736
step: 350, loss: 0.00031071799458004534
step: 360, loss: 0.000121105375001207
step: 370, loss: 4.373272895463742e-05
step: 380, loss: 4.082610394107178e-05
step: 390, loss: 9.438396227778867e-05
step: 400, loss: 3.322030897834338e-05
step: 410, loss: 0.00021203678625170141
step: 420, loss: 8.986744069261476e-05
step: 430, loss: 5.6668242905288935e-05
step: 440, loss: 0.017318855971097946
step: 450, loss: 0.008847078308463097
step: 460, loss: 3.438246130826883e-05
step: 470, loss: 0.00010709687194321305
```

```
step: 480, loss: 8.897852239897475e-05
step: 490, loss: 0.0001322790776612237
step: 500, loss: 0.00028346135513857007
step: 510, loss: 0.0004459610499907285
step: 520, loss: 0.00019068480469286442
step: 530, loss: 3.879476207657717e-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=0.9428571428571428, f1=0.9436038514442916,
best_f1=0.9429864253393664
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:16, 350.30it/s]
load_f1 = 0.9427792915531334
real_f1 = 0.9416025350837482
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, 359.21it/s]
```

4.1.4 Amazon-Google - Running the matcher

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


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

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

Selected optimization level 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: 0.8312875628471375
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06268684566020966
step: 20, loss: 0.37864312529563904
step: 30, loss: 0.3724132478237152
step: 40, loss: 0.5078404545783997
step: 50, loss: 0.2988634705543518
step: 60, loss: 0.34275156259536743
step: 70, loss: 0.2024821937084198
step: 80, loss: 0.2987602651119232
step: 90, loss: 0.3513849377632141
step: 100, loss: 0.11007294058799744
step: 110, loss: 0.3018816411495209
step: 120, loss: 0.24299943447113037
step: 130, loss: 0.19248686730861664
step: 140, loss: 0.21076005697250366
step: 150, loss: 0.27274206280708313
step: 160, loss: 0.18548907339572906
step: 170, loss: 0.09893498569726944
step: 180, loss: 0.21091267466545105
step: 190, loss: 0.11540739983320236
step: 200, loss: 0.15588781237602234
step: 210, loss: 0.4777505397796631
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.625, f1=0.6611909650924024, best_f1=0.6611909650924024
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08391958475112915
step: 10, loss: 0.041412316262722015
step: 20, loss: 0.17156384885311127
step: 30, loss: 0.2538709342479706
step: 40, loss: 0.1336417943239212
step: 50, loss: 0.2456185668706894
step: 60, loss: 0.04890062287449837
step: 70, loss: 0.14468400180339813
step: 80, loss: 0.15131419897079468
step: 90, loss: 0.08269418030977249
step: 100, loss: 0.06580799072980881
step: 110, loss: 0.0731869786977768
step: 120, loss: 0.16492892801761627
step: 130, loss: 0.15800748765468597
step: 140, loss: 0.16255410015583038
step: 150, loss: 0.07185469567775726
step: 160, loss: 0.19071367383003235
step: 170, loss: 0.1572229117155075
step: 180, loss: 0.26006564497947693
step: 190, loss: 0.10279172658920288
step: 200, loss: 0.21140789985656738
step: 210, loss: 0.21135137975215912
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6641366223908919, f1=0.7183364839319472,
best_f1=0.7183364839319472
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14637325704097748
step: 10, loss: 0.1917543262243271
step: 20, loss: 0.2748258411884308
```

```
step: 30, loss: 0.1536073088645935
step: 40, loss: 0.21726258099079132
step: 50, loss: 0.11096560209989548
step: 60, loss: 0.09414462000131607
step: 70, loss: 0.1889147013425827
step: 80, loss: 0.03827226534485817
step: 90, loss: 0.19278395175933838
step: 100, loss: 0.02223155088722706
step: 110, loss: 0.04729780554771423
step: 120, loss: 0.18568922579288483
step: 130, loss: 0.0777207463979721
step: 140, loss: 0.2741236090660095
step: 150, loss: 0.13985101878643036
step: 160, loss: 0.10463573038578033
step: 170, loss: 0.2644849419593811
step: 180, loss: 0.06554225832223892
step: 190, loss: 0.16042263805866241
step: 200, loss: 0.09648241847753525
step: 210, loss: 0.13485433161258698
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6789366053169735, f1=0.7142857142857143,
best_f1=0.7142857142857143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06336493045091629
step: 10, loss: 0.03907584398984909
step: 20, loss: 0.08045879751443863
step: 30, loss: 0.03655116260051727
step: 40, loss: 0.026650048792362213
step: 50, loss: 0.025886548683047295
step: 60, loss: 0.05400208383798599
step: 70, loss: 0.3375053405761719
step: 80, loss: 0.09178882092237473
step: 90, loss: 0.025212960317730904
step: 100, loss: 0.14703014492988586
step: 110, loss: 0.1873723715543747
step: 120, loss: 0.022359076887369156
step: 130, loss: 0.21436020731925964
step: 140, loss: 0.11910562217235565
step: 150, loss: 0.12881357967853546
step: 160, loss: 0.0976126566529274
```

```
step: 170, loss: 0.12838314473628998
step: 180, loss: 0.18978407979011536
step: 190, loss: 0.2296110838651657
step: 200, loss: 0.08053383231163025
step: 210, loss: 0.03616151958703995
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6899383983572897, f1=0.6882591093117408,
best_f1=0.6882591093117408
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.054098423570394516
step: 10, loss: 0.08590536564588547
step: 20, loss: 0.034453991800546646
step: 30, loss: 0.12505286931991577
step: 40, loss: 0.12152101844549179
step: 50, loss: 0.09762322157621384
step: 60, loss: 0.03364560753107071
step: 70, loss: 0.1069384217262268
step: 80, loss: 0.016279064118862152
step: 90, loss: 0.14309857785701752
step: 100, loss: 0.07675664871931076
step: 110, loss: 0.016054295003414154
step: 120, loss: 0.07917144149541855
step: 130, loss: 0.03449542447924614
step: 140, loss: 0.08644749969244003
step: 150, loss: 0.03655324503779411
step: 160, loss: 0.07303209602832794
step: 170, loss: 0.12970435619354248
step: 180, loss: 0.2772071659564972
step: 190, loss: 0.3987019658088684
step: 200, loss: 0.1026071310043335
step: 210, loss: 0.020494533702731133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7103174603174603, f1=0.7325581395348837,
best_f1=0.7325581395348837
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13573704659938812
step: 10, loss: 0.029183007776737213
step: 20, loss: 0.035272907465696335
step: 30, loss: 0.15261922776699066
step: 40, loss: 0.02462388575077057
step: 50, loss: 0.08525369316339493
step: 60, loss: 0.0900651142001152
step: 70, loss: 0.0054387361742556095
step: 80, loss: 0.06340061128139496
step: 90, loss: 0.09270259737968445
step: 100, loss: 0.06758511066436768
step: 110, loss: 0.10138794779777527
step: 120, loss: 0.019964979961514473
step: 130, loss: 0.0785582959651947
step: 140, loss: 0.061480455100536346
step: 150, loss: 0.014700785279273987
step: 160, loss: 0.03970171883702278
step: 170, loss: 0.03224968537688255
step: 180, loss: 0.013187071308493614
step: 190, loss: 0.13842715322971344
step: 200, loss: 0.07986854761838913
step: 210, loss: 0.05766093730926514
/usr/local/lib/python3.7/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=0.6936416184971097, f1=0.717557251908397,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026810213923454285
step: 10, loss: 0.0204398725181818
step: 20, loss: 0.049690160900354385
step: 30, loss: 0.08469630032777786
step: 40, loss: 0.03299824520945549
step: 50, loss: 0.14151611924171448
step: 60, loss: 0.1881367415189743
step: 70, loss: 0.0031249469611793756
step: 80, loss: 0.10159582644701004
step: 90, loss: 0.003976931795477867
step: 100, loss: 0.0013591052265837789
step: 110, loss: 0.14977975189685822
step: 120, loss: 0.07292363792657852
step: 130, loss: 0.015726925805211067
step: 140, loss: 0.0010798672446981072
```

```
step: 150, loss: 0.1254015415906906
step: 160, loss: 0.04610501229763031
step: 170, loss: 0.13242031633853912
step: 180, loss: 0.007708322256803513
step: 190, loss: 0.030106065794825554
step: 200, loss: 0.020478228107094765
step: 210, loss: 0.047394316643476486
/usr/local/lib/python3.7/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=0.6926229508196722, f1=0.7198364008179958,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.040483202785253525
step: 10, loss: 0.038264285773038864
step: 20, loss: 0.27436450123786926
step: 30, loss: 0.011034977622330189
step: 40, loss: 0.07922285050153732
step: 50, loss: 0.030247153714299202
step: 60, loss: 0.2203037589788437
step: 70, loss: 0.009986230172216892
step: 80, loss: 0.0687665268778801
step: 90, loss: 0.11513891071081161
step: 100, loss: 0.014210741966962814
step: 110, loss: 0.001977628795430064
step: 120, loss: 0.024466747418045998
step: 130, loss: 0.04680020362138748
step: 140, loss: 0.0043680863454937935
step: 150, loss: 0.04295551776885986
step: 160, loss: 0.15224823355674744
step: 170, loss: 0.14232134819030762
step: 180, loss: 0.03213384374976158
step: 190, loss: 0.02946183830499649
step: 200, loss: 0.037481483072042465
step: 210, loss: 0.180119127035141
/usr/local/lib/python3.7/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=0.6996197718631179, f1=0.7296296296296296,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024097973480820656
step: 10, loss: 0.11126167327165604
step: 20, loss: 0.09724889695644379
```

```
step: 30, loss: 0.1347743421792984
step: 40, loss: 0.02859606221318245
step: 50, loss: 0.008227571845054626
step: 60, loss: 0.004437049385160208
step: 70, loss: 0.026256762444972992
step: 80, loss: 0.002987263258546591
step: 90, loss: 0.060983747243881226
step: 100, loss: 0.05430399626493454
step: 110, loss: 0.02933547832071781
step: 120, loss: 0.009971930645406246
step: 130, loss: 0.006363403983414173
step: 140, loss: 0.027922436594963074
step: 150, loss: 0.019354820251464844
step: 160, loss: 0.02584953047335148
step: 170, loss: 0.008428039029240608
step: 180, loss: 0.04641718044877052
step: 190, loss: 0.07035809755325317
step: 200, loss: 0.009464451111853123
step: 210, loss: 0.06732655316591263
/usr/local/lib/python3.7/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=0.6857142857142856, f1=0.7066115702479339,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.033965352922677994
step: 10, loss: 0.0023544447030872107
step: 20, loss: 0.09300845116376877
step: 30, loss: 0.053105488419532776
step: 40, loss: 0.2390468567609787
step: 50, loss: 0.003947551362216473
step: 60, loss: 0.002018991392105818
step: 70, loss: 0.03758547827601433
step: 80, loss: 0.0022069758269935846
step: 90, loss: 0.05308815836906433
step: 100, loss: 0.002286326140165329
step: 110, loss: 0.04212208092212677
step: 120, loss: 0.013602850027382374
step: 130, loss: 0.005540247540920973
step: 140, loss: 0.0006096077268011868
step: 150, loss: 0.020618539303541183
step: 160, loss: 0.12882807850837708
step: 170, loss: 0.02671010233461857
step: 180, loss: 0.028900323435664177
step: 190, loss: 0.13616535067558289
step: 200, loss: 0.043586842715740204
```

```
step: 210, loss: 0.06843369454145432
/usr/local/lib/python3.7/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=0.7018255578093306, f1=0.7034764826175869,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019882339984178543
step: 10, loss: 0.0530741848051548
step: 20, loss: 0.00171266112010926
step: 30, loss: 0.0975276529788971
step: 40, loss: 0.01915348879992962
step: 50, loss: 0.05031118541955948
step: 60, loss: 0.04799853265285492
step: 70, loss: 0.004445101600140333
step: 80, loss: 0.1259322166442871
step: 90, loss: 0.0219937302172184
step: 100, loss: 0.06424865871667862
step: 110, loss: 0.1696425974369049
step: 120, loss: 0.07152353972196579
step: 130, loss: 0.038009416311979294
step: 140, loss: 0.14127251505851746
step: 150, loss: 0.048178497701883316
step: 160, loss: 0.003486384404823184
step: 170, loss: 0.09784442186355591
step: 180, loss: 0.01749853603541851
step: 190, loss: 0.02355470135807991
step: 200, loss: 0.00044708052882924676
step: 210, loss: 0.0022603527177125216
/usr/local/lib/python3.7/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=0.6825396825396826, f1=0.7065868263473053,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009663045406341553
step: 10, loss: 0.005921004340052605
step: 20, loss: 0.024931974709033966
step: 30, loss: 0.0020334499422460794
step: 40, loss: 0.11206911504268646
step: 50, loss: 0.09413176774978638
step: 60, loss: 0.023643091320991516
step: 70, loss: 0.0038290212396532297
step: 80, loss: 0.06833644956350327
```

```
step: 90, loss: 0.0013401606120169163
step: 100, loss: 0.0010810727253556252
step: 110, loss: 0.0014146928442642093
step: 120, loss: 0.002055032178759575
step: 130, loss: 0.009965979494154453
step: 140, loss: 0.05408618599176407
step: 150, loss: 0.0031546561513096094
step: 160, loss: 0.006114040035754442
step: 170, loss: 0.005026805214583874
step: 180, loss: 0.00673604104667902
step: 190, loss: 0.18285058438777924
step: 200, loss: 0.2267637699842453
step: 210, loss: 0.009507891722023487
/usr/local/lib/python3.7/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=0.688715953307393, f1=0.6964980544747081,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003798112738877535
step: 10, loss: 0.007048096973448992
step: 20, loss: 0.0022627655416727066
step: 30, loss: 0.10529164969921112
step: 40, loss: 0.006427429150789976
step: 50, loss: 0.00489471759647131
step: 60, loss: 0.00829872116446495
step: 70, loss: 0.16921338438987732
step: 80, loss: 0.05134349688887596
step: 90, loss: 0.040137503296136856
step: 100, loss: 0.059394072741270065
step: 110, loss: 0.007962362840771675
step: 120, loss: 0.02355242148041725
step: 130, loss: 0.01140725240111351
step: 140, loss: 0.021398356184363365
step: 150, loss: 0.0031820139847695827
step: 160, loss: 0.04213143140077591
step: 170, loss: 0.04216969758272171
step: 180, loss: 0.011670122854411602
step: 190, loss: 0.0026616742834448814
step: 200, loss: 0.007905752398073673
step: 210, loss: 0.011764760129153728
/usr/local/lib/python3.7/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=0.6845124282982792, f1=0.7007575757575758,
best_f1=0.7325581395348837
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01885189302265644
step: 10, loss: 0.014341335743665695
step: 20, loss: 0.0020923810079693794
step: 30, loss: 0.022830870002508163
step: 40, loss: 0.0051875184290111065
step: 50, loss: 0.02345287799835205
step: 60, loss: 0.05849265307188034
step: 70, loss: 0.08074356615543365
step: 80, loss: 0.04658205434679985
step: 90, loss: 0.06436672061681747
step: 100, loss: 0.06756051629781723
step: 110, loss: 0.028044115751981735
step: 120, loss: 0.030631564557552338
step: 130, loss: 0.17713142931461334
step: 140, loss: 0.0007031378336250782
step: 150, loss: 0.042433690279722214
step: 160, loss: 0.04169934615492821
step: 170, loss: 0.013661996461451054
step: 180, loss: 0.009667005389928818
step: 190, loss: 0.0022672158665955067
step: 200, loss: 0.004669326823204756
step: 210, loss: 0.008426731452345848
/usr/local/lib/python3.7/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=0.6872427983539094, f1=0.698989898989899,
best_f1=0.7325581395348837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04460152983665466
step: 10, loss: 0.04387553408741951
step: 20, loss: 0.004683818202465773
step: 30, loss: 0.015004556626081467
step: 40, loss: 0.00142187240999192
step: 50, loss: 0.0025740894488990307
step: 60, loss: 0.019927477464079857
step: 70, loss: 0.0018764882115647197
step: 80, loss: 0.018643811345100403
step: 90, loss: 0.010704726912081242
step: 100, loss: 0.058346938341856
step: 110, loss: 0.0004937078920193017
step: 120, loss: 0.06304707378149033
step: 130, loss: 0.035444412380456924
step: 140, loss: 0.004667823202908039
```

```
step: 150, loss: 0.03691258281469345
step: 160, loss: 0.010125094093382359
step: 170, loss: 0.020198682323098183
step: 180, loss: 0.04389628395438194
step: 190, loss: 0.015478438697755337
step: 200, loss: 0.03585183992981911
step: 210, loss: 0.037881508469581604
/usr/local/lib/python3.7/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=0.6918489065606361, f1=0.7045009784735813,
best_f1=0.7325581395348837
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.")
2293it [00:03, 591.82it/s]
load_f1 = 0.7070707070707072
real_f1 = 0.7034764826175869
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, 361.08it/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 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
```

407

```
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: 0.8497321605682373
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1673612892627716
step: 20, loss: 0.15069308876991272
step: 30, loss: 0.5104325413703918
step: 40, loss: 0.2618201673030853
step: 50, loss: 0.31128475069999695
step: 60, loss: 0.3670501410961151
step: 70, loss: 0.1802954226732254
step: 80, loss: 0.5449020862579346
step: 90, loss: 0.2508680522441864
step: 100, loss: 0.21679645776748657
step: 110, loss: 0.2378239780664444
step: 120, loss: 0.4123939573764801
step: 130, loss: 0.346694678068161
step: 140, loss: 0.30115601420402527
step: 150, loss: 0.26629310846328735
step: 160, loss: 0.22845685482025146
step: 170, loss: 0.3424012362957001
step: 180, loss: 0.2749902606010437
```

```
step: 190, loss: 0.14878065884113312
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6437994722955146, f1=0.6410958904109589,
best_f1=0.6410958904109589
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24572165310382843
step: 10, loss: 0.050814609974622726
step: 20, loss: 0.09447502344846725
step: 30, loss: 0.12571528553962708
step: 40, loss: 0.5907012224197388
step: 50, loss: 0.2835958003997803
step: 60, loss: 0.15490776300430298
step: 70, loss: 0.18163755536079407
step: 80, loss: 0.1359197050333023
step: 90, loss: 0.12251687794923782
step: 100, loss: 0.2676180899143219
step: 110, loss: 0.07494264841079712
step: 120, loss: 0.13271458446979523
step: 130, loss: 0.16130298376083374
step: 140, loss: 0.14352752268314362
step: 150, loss: 0.05362597852945328
step: 160, loss: 0.11010529845952988
step: 170, loss: 0.14308969676494598
step: 180, loss: 0.10308484733104706
step: 190, loss: 0.12354882061481476
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7539267015706806, f1=0.7731958762886598,
best_f1=0.7731958762886598
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12140709906816483
step: 10, loss: 0.20287472009658813
step: 20, loss: 0.12531307339668274
```

```
step: 30, loss: 0.03990066424012184
step: 40, loss: 0.048714492470026016
step: 50, loss: 0.1754387617111206
step: 60, loss: 0.05050548166036606
step: 70, loss: 0.06712624430656433
step: 80, loss: 0.4342673420906067
step: 90, loss: 0.03661174699664116
step: 100, loss: 0.12667836248874664
step: 110, loss: 0.0843164324760437
step: 120, loss: 0.015831297263503075
step: 130, loss: 0.07183688879013062
step: 140, loss: 0.03164132684469223
step: 150, loss: 0.08683522790670395
step: 160, loss: 0.13887102901935577
step: 170, loss: 0.12978537380695343
step: 180, loss: 0.10358330607414246
step: 190, loss: 0.07226449996232986
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7560321715817695, f1=0.7578947368421052,
best_f1=0.7578947368421052
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23905709385871887
step: 10, loss: 0.023924799636006355
step: 20, loss: 0.03201677277684212
step: 30, loss: 0.04103092849254608
step: 40, loss: 0.0036278278566896915
step: 50, loss: 0.01193210855126381
step: 60, loss: 0.203238382935524
step: 70, loss: 0.01782291755080223
step: 80, loss: 0.016160091385245323
step: 90, loss: 0.02001318335533142
step: 100, loss: 0.010265922173857689
step: 110, loss: 0.020376110449433327
step: 120, loss: 0.05092398077249527
step: 130, loss: 0.22664505243301392
step: 140, loss: 0.0797068327665329
step: 150, loss: 0.05997241288423538
step: 160, loss: 0.08039435744285583
step: 170, loss: 0.013230529613792896
step: 180, loss: 0.09470947831869125
```

```
step: 190, loss: 0.07721846550703049
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8184281842818427, f1=0.7851458885941645,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09591509401798248
step: 10, loss: 0.010571437887847424
step: 20, loss: 0.06811273843050003
step: 30, loss: 0.037559282034635544
step: 40, loss: 0.013092370703816414
step: 50, loss: 0.009018032811582088
step: 60, loss: 0.005608661565929651
step: 70, loss: 0.008617633022367954
step: 80, loss: 0.005225652363151312
step: 90, loss: 0.0031332450453191996
step: 100, loss: 0.039046794176101685
step: 110, loss: 0.018964391201734543
step: 120, loss: 0.0017390715656802058
step: 130, loss: 0.00455763703212142
step: 140, loss: 0.005104129668325186
step: 150, loss: 0.08425799012184143
step: 160, loss: 0.05990096926689148
step: 170, loss: 0.025042643770575523
step: 180, loss: 0.03801334276795387
step: 190, loss: 0.08921369910240173
/usr/local/lib/python3.7/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=0.7826086956521738, f1=0.7771739130434783,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0808653011918068
step: 10, loss: 0.0006108502275310457
step: 20, loss: 0.01783432625234127
step: 30, loss: 0.0009635729948058724
step: 40, loss: 0.024983402341604233
step: 50, loss: 0.004378859885036945
step: 60, loss: 0.0037874553818255663
```

```
step: 70, loss: 0.17548398673534393
step: 80, loss: 0.0510752871632576
step: 90, loss: 0.01791069097816944
step: 100, loss: 0.029073482379317284
step: 110, loss: 0.08331792801618576
step: 120, loss: 0.017411954700946808
step: 130, loss: 0.003932386636734009
step: 140, loss: 0.0035644788295030594
step: 150, loss: 0.05033827945590019
step: 160, loss: 0.0037227943539619446
step: 170, loss: 0.026810547336935997
step: 180, loss: 0.008113658986985683
step: 190, loss: 0.003478520317003131
/usr/local/lib/python3.7/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=0.7826086956521738, f1=0.8010899182561309,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005077355541288853
step: 10, loss: 0.09582503139972687
step: 20, loss: 0.0027805063873529434
step: 30, loss: 0.013935108669102192
step: 40, loss: 0.0020349803380668163
step: 50, loss: 0.08823015540838242
step: 60, loss: 0.0031268501188606024
step: 70, loss: 0.0017045763088390231
step: 80, loss: 0.18160918354988098
step: 90, loss: 0.12121982872486115
step: 100, loss: 0.01852356642484665
step: 110, loss: 0.0036109606735408306
step: 120, loss: 0.0393357090651989
step: 130, loss: 0.001564580830745399
step: 140, loss: 0.0014803396770730615
step: 150, loss: 0.06925881654024124
step: 160, loss: 0.0664459615945816
step: 170, loss: 0.0032144798897206783
step: 180, loss: 0.011079424992203712
step: 190, loss: 0.014323897659778595
/usr/local/lib/python3.7/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=0.7786666666666667, f1=0.7905759162303665,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000750982144381851
step: 10, loss: 0.005214715376496315
step: 20, loss: 0.0005967431352473795
step: 30, loss: 0.001448154216632247
step: 40, loss: 0.0035243318416178226
step: 50, loss: 0.026874661445617676
step: 60, loss: 0.002383469371125102
step: 70, loss: 0.0008774145389907062
step: 80, loss: 0.1887253224849701
step: 90, loss: 0.016654737293720245
step: 100, loss: 0.03417243808507919
step: 110, loss: 0.00901323463767767
step: 120, loss: 0.013656949624419212
step: 130, loss: 0.013834532350301743
step: 140, loss: 0.0032198845874518156
step: 150, loss: 0.0027204041834920645
step: 160, loss: 0.005214049946516752
step: 170, loss: 0.0004190432373434305
step: 180, loss: 0.06225904822349548
step: 190, loss: 0.03917583078145981
/usr/local/lib/python3.7/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=0.7925531914893618, f1=0.8032345013477089,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0059703947044909
step: 10, loss: 0.004587628412991762
step: 20, loss: 0.010970494709908962
step: 30, loss: 0.0008603947353549302
step: 40, loss: 0.002098072087392211
step: 50, loss: 0.0030430445913225412
step: 60, loss: 0.001183483051136136
step: 70, loss: 0.005725252442061901
step: 80, loss: 0.005441715009510517
step: 90, loss: 0.0028769196942448616
step: 100, loss: 0.0007913283770903945
step: 110, loss: 0.002470972714945674
step: 120, loss: 0.002264046110212803
step: 130, loss: 0.008824196644127369
step: 140, loss: 0.005543767474591732
step: 150, loss: 0.009392053820192814
step: 160, loss: 0.11876217275857925
step: 170, loss: 0.004059627186506987
step: 180, loss: 0.01481183897703886
```

step: 190, loss: 0.023369023576378822
/usr/local/lib/python3.7/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=0.7926509186351706, f1=0.7745358090185677,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000224187460844405
step: 10, loss: 0.00048991142659269273
step: 20, loss: 0.0005885956925339997
step: 30, loss: 0.0021198487374931574
step: 40, loss: 0.0005559954443015158
step: 50, loss: 0.0011930213076993823
step: 60, loss: 0.0011798305520413816
step: 70, loss: 0.0009691455634310842
step: 80, loss: 0.010828044265508652
step: 90, loss: 0.0024494221433997154
step: 100, loss: 0.0014843742828816175
step: 110, loss: 0.008644355461001396
step: 120, loss: 0.0013142158277332783
step: 130, loss: 0.17667661607265472
step: 140, loss: 0.04476344957947731
step: 150, loss: 0.0014946766896173358
step: 160, loss: 0.0006368040922097862
step: 170, loss: 0.009644618257880211
step: 180, loss: 0.0022249144967645407
step: 190, loss: 0.03182763233780861
/usr/local/lib/python3.7/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=0.7967914438502673, f1=0.8106666666666666,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005302668432705104
step: 10, loss: 0.00039193700649775565
step: 20, loss: 0.0007120580412447453
step: 30, loss: 0.00896934513002634
step: 40, loss: 0.0006717179785482585
step: 50, loss: 0.0010457327589392662
step: 60, loss: 0.0004785367927979678
step: 70, loss: 0.00020245887571945786
step: 80, loss: 0.00016335661348421127
step: 90, loss: 0.06774646043777466
step: 100, loss: 0.0009838053956627846

```
step: 110, loss: 0.003311609150841832
step: 120, loss: 0.09179321676492691
step: 130, loss: 0.0009377130772918463
step: 140, loss: 0.0008470265311188996
step: 150, loss: 0.0043282113038003445
step: 160, loss: 0.05229324847459793
step: 170, loss: 0.003940979018807411
step: 180, loss: 0.04327864944934845
step: 190, loss: 0.04699026793241501
/usr/local/lib/python3.7/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=0.7904509283819628, f1=0.7946666666666666,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004804382566362619
step: 10, loss: 0.006681989412754774
step: 20, loss: 0.000957260315772146
step: 30, loss: 0.01578815095126629
step: 40, loss: 0.032989151775836945
step: 50, loss: 0.016198046505451202
step: 60, loss: 0.046418581157922745
step: 70, loss: 0.005666413344442844
step: 80, loss: 0.003290038090199232
step: 90, loss: 0.005065475590527058
step: 100, loss: 0.002397479023784399
step: 110, loss: 0.0038239231798797846
step: 120, loss: 0.0004606003640219569
step: 130, loss: 0.0009190443670377135
step: 140, loss: 0.03605847433209419
step: 150, loss: 0.00047682964941486716
step: 160, loss: 0.0036296811886131763
step: 170, loss: 0.0006180735654197633
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 180, loss: 0.0033825368154793978
step: 190, loss: 0.05637086182832718
/usr/local/lib/python3.7/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=0.7883597883597884, f1=0.7989417989417988,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005868873093277216
step: 10, loss: 0.0033973357640206814
```

```
step: 20, loss: 0.010785103775560856
step: 30, loss: 0.001958470093086362
step: 40, loss: 0.0015522617613896728
step: 50, loss: 0.020804861560463905
step: 60, loss: 0.0019525891402736306
step: 70, loss: 0.045118298381567
step: 80, loss: 0.0003356946399435401
step: 90, loss: 0.013682444579899311
step: 100, loss: 0.00044975016498938203
step: 110, loss: 0.0004297529230825603
step: 120, loss: 0.0006569231045432389
step: 130, loss: 0.02445036917924881
step: 140, loss: 0.0009403575677424669
step: 150, loss: 0.003288260428234935
step: 160, loss: 0.0012810304760932922
step: 170, loss: 0.00026577452081255615
step: 180, loss: 0.0011014725314453244
step: 190, loss: 0.000954379967879504
/usr/local/lib/python3.7/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=0.7896103896103895, f1=0.7967914438502673,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010075510945171118
step: 10, loss: 0.025137823075056076
step: 20, loss: 0.00072482996620223783
step: 30, loss: 0.0016423083143308759
step: 40, loss: 0.0002838324289768934
step: 50, loss: 0.0007386454381048679
step: 60, loss: 0.0009290669113397598
step: 70, loss: 0.0015582398045808077
step: 80, loss: 0.00022550616995431483
step: 90, loss: 0.005691688973456621
step: 100, loss: 0.0020939514506608248
step: 110, loss: 0.0013314802199602127
step: 120, loss: 0.0010316764237359166
step: 130, loss: 0.0010076704202219844
step: 140, loss: 0.0009758798987604678
step: 150, loss: 0.0008872828329913318
step: 160, loss: 0.040578510612249374
step: 170, loss: 0.012330849654972553
step: 180, loss: 0.0011936596129089594
step: 190, loss: 0.0001803345512598753
/usr/local/lib/python3.7/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=0.7866323907455013, f1=0.7926509186351706,
best_f1=0.7851458885941645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00046939734602347016
step: 10, loss: 0.03977908939123154
step: 20, loss: 0.00048471850459463894
step: 30, loss: 0.058883361518383026
step: 40, loss: 0.0006637248443439603
step: 50, loss: 0.0027241960633546114
step: 60, loss: 0.0041723474860191345
step: 70, loss: 0.0012225533137097955
step: 80, loss: 0.0009412406361661851
step: 90, loss: 0.000829110445920378
step: 100, loss: 0.0002971617389693856
step: 110, loss: 0.0010411859257146716
step: 120, loss: 0.009401076473295689
step: 130, loss: 0.0010229343315586448
step: 140, loss: 0.0020936201326549053
step: 150, loss: 0.0005190664087422192
step: 160, loss: 0.00047963616088964045
step: 170, loss: 0.00851002149283886
step: 180, loss: 0.0007426484953612089
step: 190, loss: 0.00048798121861182153
/usr/local/lib/python3.7/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=0.7968337730870713, f1=0.7904509283819628,
best_f1=0.7851458885941645
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 [00:06, 295.58it/s]
load_f1 = 0.6432432432432432
real_f1 = 0.6288659793814433
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, 362.98it/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
```

```python
# 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: 0.8407676815986633
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.2255047857761383
step: 20, loss: 0.1524595022201538
step: 30, loss: 0.23942196369171143
step: 40, loss: 0.31467941403388977
step: 50, loss: 0.3728472888469696
step: 60, loss: 0.4450845420360565
step: 70, loss: 0.31017616391181946
step: 80, loss: 0.2572568356990814
step: 90, loss: 0.40054047107696533
step: 100, loss: 0.22813032567501068
step: 110, loss: 0.17676453292369843
step: 120, loss: 0.5466529726982117
step: 130, loss: 0.42512333393096924
step: 140, loss: 0.47291845083236694
step: 150, loss: 0.11120130866765976
step: 160, loss: 0.3148733973503113
step: 170, loss: 0.1882244050502777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6338028169014085, f1=0.6431924882629109,
best_f1=0.6431924882629109
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20901274681091309
step: 10, loss: 0.06208238750696182
step: 20, loss: 0.18515852093696594
step: 30, loss: 0.16226662695407867
step: 40, loss: 0.1338624507188797
step: 50, loss: 0.25446730852127075
step: 60, loss: 0.053934428840875626
step: 70, loss: 0.22835169732570648
step: 80, loss: 0.19018599390983582
step: 90, loss: 0.10423773527145386
step: 100, loss: 0.09880726784467697
step: 110, loss: 0.19889236986637115
step: 120, loss: 0.05637515336275101
step: 130, loss: 0.0649668425321579
step: 140, loss: 0.0989861860871315
```

```
step: 150, loss: 0.09332218766212463
step: 160, loss: 0.1274566948413849
step: 170, loss: 0.1031746044754982
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7885117493472584, f1=0.7676767676767677,
best_f1=0.7676767676767677
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03321788087487221
step: 10, loss: 0.0400581918656826
step: 20, loss: 0.010347127914428711
step: 30, loss: 0.1546122133731842
step: 40, loss: 0.020254800096154213
step: 50, loss: 0.15831667184829712
step: 60, loss: 0.14100293815135956
step: 70, loss: 0.09729170054197311
step: 80, loss: 0.06122240424156189
step: 90, loss: 0.07676637172698975
step: 100, loss: 0.03136889636516571
step: 110, loss: 0.07542166113853455
step: 120, loss: 0.006965682841837406
step: 130, loss: 0.12004747241735458
step: 140, loss: 0.005937815178185701
step: 150, loss: 0.016728214919567108
step: 160, loss: 0.15459682047367096
step: 170, loss: 0.10484650731086731
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7903614457831325, f1=0.7857142857142857,
best_f1=0.7857142857142857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.039015140384435654
step: 10, loss: 0.2594696283340454
step: 20, loss: 0.02096734382212162
```

```
step: 30, loss: 0.09209642559289932
step: 40, loss: 0.15622079372406006
step: 50, loss: 0.00952246505767107
step: 60, loss: 0.12147263437509537
step: 70, loss: 0.02135992981493473
step: 80, loss: 0.15217925608158112
step: 90, loss: 0.013455861248075962
step: 100, loss: 0.019786091521382332
step: 110, loss: 0.019644711166620255
step: 120, loss: 0.13187673687934875
step: 130, loss: 0.06138479709625244
step: 140, loss: 0.023404879495501518
step: 150, loss: 0.022307848557829857
step: 160, loss: 0.11343089491128922
step: 170, loss: 0.005159557331353426
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8160377358490567, f1=0.8082191780821918,
best_f1=0.8082191780821918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03931188955903053
step: 10, loss: 0.053200289607048035
step: 20, loss: 0.00521688349545002
step: 30, loss: 0.05241430923342705
step: 40, loss: 0.008515291847288609
step: 50, loss: 0.07489732652902603
step: 60, loss: 0.0026468709111213684
step: 70, loss: 0.015618595294465437
step: 80, loss: 0.006268445868045092
step: 90, loss: 0.05358852446079254
step: 100, loss: 0.03601501137018204
step: 110, loss: 0.13357700407505035
step: 120, loss: 0.06825301051139832
step: 130, loss: 0.045190636068582535
step: 140, loss: 0.038254477083683014
step: 150, loss: 0.007929375395178795
step: 160, loss: 0.009130241349339485
step: 170, loss: 0.0597820058465004
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8320413436692506, f1=0.8418491484184915,
best_f1=0.8418491484184915
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14161472022533417
step: 10, loss: 0.051117170602083206
step: 20, loss: 0.014489213936030865
step: 30, loss: 0.005053515080362558
step: 40, loss: 0.015822557732462883
step: 50, loss: 0.026064341887831688
step: 60, loss: 0.031196558848023415
step: 70, loss: 0.01832989789545536
step: 80, loss: 0.026731083169579506
step: 90, loss: 0.09896570444107056
step: 100, loss: 0.03514060750603676
step: 110, loss: 0.025025201961398125
step: 120, loss: 0.021935727447271347
step: 130, loss: 0.1717178225517273
step: 140, loss: 0.020794577896595
step: 150, loss: 0.12416130304336548
step: 160, loss: 0.04876509681344032
step: 170, loss: 0.09967125207185745
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8477157360406091, f1=0.8446601941747571,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0031835436820983887
step: 10, loss: 0.001226787338964641
step: 20, loss: 0.010499412193894386
step: 30, loss: 0.0044127097353339195
step: 40, loss: 0.017387932166457176
step: 50, loss: 0.00960177555680275
step: 60, loss: 0.1111830323934555
step: 70, loss: 0.007626068312674761
step: 80, loss: 0.002375954994931817

```
step: 90, loss: 0.014916780404746532
step: 100, loss: 0.006620299071073532
step: 110, loss: 0.0033070542849600315
step: 120, loss: 0.12320712208747864
step: 130, loss: 0.07655705511569977
step: 140, loss: 0.0006771271000616252
step: 150, loss: 0.019847005605697632
step: 160, loss: 0.0041633895598351955
step: 170, loss: 0.018435189500451088
/usr/local/lib/python3.7/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=0.8402948402948404, f1=0.8441247002398082,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010560248047113419
step: 10, loss: 0.010413400828838348
step: 20, loss: 0.0010658020619302988
step: 30, loss: 0.0023294580169022083
step: 40, loss: 0.0032021929509937763
step: 50, loss: 0.0037748629692941904
step: 60, loss: 0.005935401190072298
step: 70, loss: 0.025763923302292824
step: 80, loss: 0.0020186000037938356
step: 90, loss: 0.006271291524171829
step: 100, loss: 0.0035208603367209435
step: 110, loss: 0.06249464675784111
step: 120, loss: 0.0032048732973635197
step: 130, loss: 0.0028762726578861475
step: 140, loss: 0.035609785467386246
step: 150, loss: 0.20885278284549713
step: 160, loss: 0.09900493174791336
step: 170, loss: 0.011254806071519852
/usr/local/lib/python3.7/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=0.8243902439024391, f1=0.8364485981308412,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04176346957683563
step: 10, loss: 0.005586151499301195
step: 20, loss: 0.038706790655851364
step: 30, loss: 0.01650385558605194
step: 40, loss: 0.011778153479099274
```

```
step: 50, loss: 0.021130265668034554
step: 60, loss: 0.0012352167395874858
step: 70, loss: 0.009192313067615032
step: 80, loss: 0.036652132868766785
step: 90, loss: 0.029951991513371468
step: 100, loss: 0.044018588960170746
step: 110, loss: 0.03247058764100075
step: 120, loss: 0.008942296728491783
step: 130, loss: 0.0010775609407573938
step: 140, loss: 0.002071174792945385
step: 150, loss: 0.0150986947119236
step: 160, loss: 0.06530686467885971
step: 170, loss: 0.009609154425561428
/usr/local/lib/python3.7/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=0.8337468982630273, f1=0.8402948402948404,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001061749178916216
step: 10, loss: 0.06070808321237564
step: 20, loss: 0.0015410479390993714
step: 30, loss: 0.0004599982057698071
step: 40, loss: 0.02373790740966797
step: 50, loss: 0.00855984352529049
step: 60, loss: 0.002079092198982835
step: 70, loss: 0.0039213672280311584
step: 80, loss: 0.002187928417697549
step: 90, loss: 0.045416459441185
step: 100, loss: 0.000267598225036636
step: 110, loss: 0.011148408055305481
step: 120, loss: 0.0800342783331871
step: 130, loss: 0.009728792123496532
step: 140, loss: 0.010763094760477543
step: 150, loss: 0.001887128921225667
step: 160, loss: 0.0005361801013350487
step: 170, loss: 0.001526552950963378
/usr/local/lib/python3.7/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=0.8358974358974359, f1=0.85, best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033227092353627086
step: 10, loss: 0.008055376820266247
```

```
step: 20, loss: 0.008490580134093761
step: 30, loss: 0.004027340095490217
step: 40, loss: 0.002830411307513714
step: 50, loss: 0.00020221018348820508
step: 60, loss: 0.0019016268197447062
step: 70, loss: 0.026198552921414375
step: 80, loss: 0.00036358486977405846
step: 90, loss: 0.0031356438994407654
step: 100, loss: 0.0005056296358816326
step: 110, loss: 0.01971130445599556
step: 120, loss: 0.0576251782476902
step: 130, loss: 0.007607117760926485
step: 140, loss: 0.02672976814210415
step: 150, loss: 0.003199234139174223
step: 160, loss: 0.0004719808930531144
step: 170, loss: 0.09534341096878052
/usr/local/lib/python3.7/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=0.8423645320197045, f1=0.8476190476190476,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028060803189873695
step: 10, loss: 0.011708437465131283
step: 20, loss: 0.027552397921681404
step: 30, loss: 0.001412300392985344
step: 40, loss: 0.002183504868298769
step: 50, loss: 0.0007475687307305634
step: 60, loss: 0.0010006324155256152
step: 70, loss: 0.019207224249839783
step: 80, loss: 0.0021406731102615595
step: 90, loss: 0.0006960509927012026
step: 100, loss: 0.001119117601774633
step: 110, loss: 0.0003416786785237491
step: 120, loss: 0.008567187935113907
step: 130, loss: 0.003181903623044491
step: 140, loss: 0.006421452853828669
step: 150, loss: 0.0024446535389870405
step: 160, loss: 0.005466947332024574
step: 170, loss: 0.0008282887865789235
/usr/local/lib/python3.7/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=0.8415584415584415, f1=0.8556962025316455,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007234509685076773
step: 10, loss: 0.061616405844688416
step: 20, loss: 0.005418387241661549
step: 30, loss: 0.0007100724615156651
step: 40, loss: 0.004387456923723221
step: 50, loss: 0.00398282753303647
step: 60, loss: 0.0021946753840893507
step: 70, loss: 0.010226658545434475
step: 80, loss: 0.0037853543180972338
step: 90, loss: 0.00025341782020404935
step: 100, loss: 0.0005244897329248488
step: 110, loss: 0.00530729815363884
step: 120, loss: 0.001636928296647966
step: 130, loss: 0.011986812576651573
step: 140, loss: 0.0006529943784698844
step: 150, loss: 0.0013471078127622604
step: 160, loss: 0.007247992791235447
step: 170, loss: 0.0032281088642776012
/usr/local/lib/python3.7/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=0.8390501319261214, f1=0.8535353535353535,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006181205972097814
step: 10, loss: 0.0010672982316464186
step: 20, loss: 0.008708113804459572
step: 30, loss: 0.006665349937975407
step: 40, loss: 0.0005523552536033094
step: 50, loss: 0.014429259113967419
step: 60, loss: 0.00015624794468749315
step: 70, loss: 0.025671498849987984
step: 80, loss: 0.017952919006347656
step: 90, loss: 0.0008407753775827587
step: 100, loss: 0.00023408683773359584
step: 110, loss: 0.005563647020608187
step: 120, loss: 0.001150817726738751
step: 130, loss: 0.0005608173087239265
step: 140, loss: 0.0037817771080881357
step: 150, loss: 0.00034307388705201447
step: 160, loss: 0.04332312196493149
step: 170, loss: 0.0020091969054192305
/usr/local/lib/python3.7/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=0.8472906403940887,
best_f1=0.8446601941747571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019462943309918046
step: 10, loss: 0.017840251326560974
step: 20, loss: 0.013689895160496235
step: 30, loss: 0.07922019809484482
step: 40, loss: 0.0010505809914320707
step: 50, loss: 0.0037234032060950994
step: 60, loss: 0.0003555998264346272
step: 70, loss: 0.0011414249893277884
step: 80, loss: 0.0015050637302920222
step: 90, loss: 0.00977443065494299
step: 100, loss: 0.0007751249358989298
step: 110, loss: 0.0002468783059157431
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 120, loss: 0.01231394149363041
step: 130, loss: 0.0007691342034377158
step: 140, loss: 0.0012599336914718151
step: 150, loss: 0.04690517485141754
step: 160, loss: 0.00018202503270003945
step: 170, loss: 0.002573601668700576
/usr/local/lib/python3.7/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=0.8375634517766498, f1=0.8472906403940887,
best_f1=0.8446601941747571
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, 368.09it/s]
load_f1 = 0.4541484716157206
real_f1 = 0.41045751633986927
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, 345.44it/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: 0.8051916360855103
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.48251381516456604
step: 20, loss: 0.6021324992179871
step: 30, loss: 0.432443231344223
step: 40, loss: 0.22934381663799286
step: 50, loss: 0.10548994690179825
step: 60, loss: 0.0811690017580986
step: 70, loss: 0.0869748592376709
step: 80, loss: 0.2584010660648346
step: 90, loss: 0.04056978225708008
step: 100, loss: 0.053617995232343674
step: 110, loss: 0.07524780929088593
step: 120, loss: 0.08901704847812653
step: 130, loss: 0.014605945907533169
step: 140, loss: 0.0705818384885788
step: 150, loss: 0.11784327775239944
step: 160, loss: 0.2589518427848816
step: 170, loss: 0.05774664506316185
step: 180, loss: 0.019537469372153282
step: 190, loss: 0.14421771466732025
step: 200, loss: 0.012691622599959373
step: 210, loss: 0.11981730908155441
step: 220, loss: 0.029042953625321388
step: 230, loss: 0.013676974922418594
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9708520179372198, f1=0.972972972972973,
best_f1=0.972972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.064497413665056229
step: 10, loss: 0.08544980734586716
step: 20, loss: 0.004783228505402803
step: 30, loss: 0.003281492507085204
step: 40, loss: 0.006651859264820814
step: 50, loss: 0.017010098323225975
step: 60, loss: 0.005377314053475857
step: 70, loss: 0.04370128735899925
step: 80, loss: 0.031311314553022385
step: 90, loss: 0.006749424152076244
```

```
step: 100, loss: 0.1358826756477356
step: 110, loss: 0.22398939728736877
step: 120, loss: 0.004313752520829439
step: 130, loss: 0.008606290444731712
step: 140, loss: 0.2803487777709961
step: 150, loss: 0.02126733399927616
step: 160, loss: 0.00856422446668148
step: 170, loss: 0.04514235630631447
step: 180, loss: 0.016408730298280716
step: 190, loss: 0.10030826181173325
step: 200, loss: 0.009835186414420605
step: 210, loss: 0.06361591815948486
step: 220, loss: 0.001508713117800653
step: 230, loss: 0.006685391999781132
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9720670391061451, f1=0.9753914988814317,
best_f1=0.9753914988814317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.047235406935214996
step: 10, loss: 0.031400181353092194
step: 20, loss: 0.0031593532767146826
step: 30, loss: 0.007680639624595642
step: 40, loss: 0.007872957736253738
step: 50, loss: 0.009655366651713848
step: 60, loss: 0.06245177984237671
step: 70, loss: 0.004408303648233414
step: 80, loss: 0.005806189030408859
step: 90, loss: 0.0037515393923968077
step: 100, loss: 0.0017384585225954652
step: 110, loss: 0.006708959117531776
step: 120, loss: 0.013519949279725552
step: 130, loss: 0.004775695502758026
step: 140, loss: 0.0029221312142908573
step: 150, loss: 0.0023627416230738163
step: 160, loss: 0.006866587325930595
step: 170, loss: 0.09627865999937057
step: 180, loss: 0.01611579768359661
step: 190, loss: 0.03164180740714073
step: 200, loss: 0.0046662939712405205
step: 210, loss: 0.026481401175260544
```

step: 220, loss: 0.003790762508288026
step: 230, loss: 0.07956629991531372
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9787709497206705, 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: 0.022951563820242882
step: 10, loss: 0.005201323423534632
step: 20, loss: 0.03995215520262718
step: 30, loss: 0.0031971510034054518
step: 40, loss: 0.030026735737919807
step: 50, loss: 0.003322877688333392
step: 60, loss: 0.0066225831396877766
step: 70, loss: 0.15343588590621948
step: 80, loss: 0.13670696318149567
step: 90, loss: 0.010939819738268852
step: 100, loss: 0.000826436560600996
step: 110, loss: 0.06744597852230072
step: 120, loss: 0.0735960528254509
step: 130, loss: 0.004100088495761156
step: 140, loss: 0.00036262947833165526
step: 150, loss: 0.0010144540574401617
step: 160, loss: 0.0010916200699284673
step: 170, loss: 0.022587308660149574
step: 180, loss: 0.07578085362911224
step: 190, loss: 0.004732104949653149
step: 200, loss: 0.008218125440180302
step: 210, loss: 0.0011455257190391421
step: 220, loss: 0.0010503832018002868
step: 230, loss: 0.0008336245082318783
/usr/local/lib/python3.7/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=0.9733333333333333, f1=0.9711111111111111,
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: 0.0029571019113063812
step: 10, loss: 0.0023914065677672625

```
step: 20, loss: 0.0011001405073329806
step: 30, loss: 0.0009964596247300506
step: 40, loss: 0.000586148293223232
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.00639126542955637
step: 60, loss: 0.0006900850567035377
step: 70, loss: 0.010112639516592026
step: 80, loss: 0.17808832228183746
step: 90, loss: 0.04739941656589508
step: 100, loss: 0.006001878529787064
step: 110, loss: 0.010958199389278889
step: 120, loss: 0.0516454242169857
step: 130, loss: 0.007926229387521744
step: 140, loss: 0.0012483749305829406
step: 150, loss: 0.0005656600114889443
step: 160, loss: 0.10298389196395874
step: 170, loss: 0.02889932319521904
step: 180, loss: 0.0015782004920765758
step: 190, loss: 0.0074719819240272045
step: 200, loss: 0.0215896125882864
step: 210, loss: 0.003417547792196274
step: 220, loss: 0.005539043806493282
step: 230, loss: 0.007943084463477135
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.978912319644839, 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: 0.007442818023264408
step: 10, loss: 0.0015242531662806869
step: 20, loss: 0.0019263217691332102
step: 30, loss: 0.002729538595303893
step: 40, loss: 0.0023482569959014654
step: 50, loss: 0.036531686782836914
step: 60, loss: 0.019116194918751717
step: 70, loss: 0.0064058247953653336
step: 80, loss: 0.002273701364174485
step: 90, loss: 0.0012215763563290238
step: 100, loss: 0.0034212705213576555
step: 110, loss: 0.038460198789834976
step: 120, loss: 0.0016527456464245915
```

```
step: 130, loss: 0.0036333876196295023
step: 140, loss: 0.0054648988880217075
step: 150, loss: 0.0015580322360619903
step: 160, loss: 0.00258809642121196
step: 170, loss: 0.000911062874365598
step: 180, loss: 0.0007838410092517734
step: 190, loss: 0.013272110372781754
step: 200, loss: 0.0034507219679653645
step: 210, loss: 0.002706265775486827
step: 220, loss: 0.00031097716419026256
step: 230, loss: 0.0014972459757700562
/usr/local/lib/python3.7/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=0.9741863075196409, f1=0.9764309764309763,
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: 0.09067334979772568
step: 10, loss: 0.0015339527744799852
step: 20, loss: 0.0009868997149169445
step: 30, loss: 0.0007003210484981537
step: 40, loss: 0.004863186273723841
step: 50, loss: 0.00030630864785052836
step: 60, loss: 0.012861398980021477
step: 70, loss: 0.0016683520516380668
step: 80, loss: 0.00031139643397182226
step: 90, loss: 0.007583972532302141
step: 100, loss: 0.0013875886797904968
step: 110, loss: 0.013603700324892998
step: 120, loss: 0.0025019419845193624
step: 130, loss: 0.010942612774670124
step: 140, loss: 0.00012327164586167783
step: 150, loss: 0.005513571668416262
step: 160, loss: 0.00012545475328806788
step: 170, loss: 0.002431404311209917
step: 180, loss: 0.00019630031601991504
step: 190, loss: 0.0028877269942313433
step: 200, loss: 0.010885702446103096
step: 210, loss: 0.0008364195236936212
step: 220, loss: 0.004203852731734514
step: 230, loss: 0.024618929252028465
/usr/local/lib/python3.7/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=0.9776286353467561, f1=0.9742441209406495,
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: 0.06451524794101715
step: 10, loss: 0.008898700587451458
step: 20, loss: 0.00012109615636290982
step: 30, loss: 0.0017965106526389718
step: 40, loss: 0.00052652362501240872
step: 50, loss: 0.0029242150485515594
step: 60, loss: 0.00017667746578808874
step: 70, loss: 0.00016377554857172072
step: 80, loss: 0.0015142472693696618
step: 90, loss: 0.0001575346541358158
step: 100, loss: 0.0007889169501140714
step: 110, loss: 0.0001717969571473077
step: 120, loss: 8.828224235912785e-05
step: 130, loss: 0.011422966606914997
step: 140, loss: 0.0001549370208522305
step: 150, loss: 8.155364048434421e-05
step: 160, loss: 0.00014968644245527685
step: 170, loss: 9.907486673910171e-05
step: 180, loss: 0.0006704877596348524
step: 190, loss: 0.0003444056201260537
step: 200, loss: 0.05999429151415825
step: 210, loss: 0.0007230945629999042
step: 220, loss: 0.0023948170710355043
step: 230, loss: 0.05770060047507286
/usr/local/lib/python3.7/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=0.9754464285714286, f1=0.9744160177975528,
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: 0.00011162098235217854
step: 10, loss: 0.0013617553049698472
step: 20, loss: 0.0014425149420276284
step: 30, loss: 0.0001807007211027667
step: 40, loss: 6.20175851508975e-05
step: 50, loss: 0.0420515239238739
step: 60, loss: 0.0021317540667951107
step: 70, loss: 0.0007449687691405416
step: 80, loss: 0.03189615160226822
step: 90, loss: 0.007117887027561665
step: 100, loss: 0.0026566984597593546
step: 110, loss: 0.0010777104180306196
step: 120, loss: 0.03649887070059776
```

```
step: 130, loss: 7.952156738610938e-05
step: 140, loss: 0.017275338992476463
step: 150, loss: 6.702127575408667e-05
step: 160, loss: 0.0002165732003049925
step: 170, loss: 0.0001628839672775939
step: 180, loss: 0.0051561398431658745
step: 190, loss: 0.0002100663841702044
step: 200, loss: 0.0009343436686322093
step: 210, loss: 0.0029416207689791918
step: 220, loss: 0.029451079666614532
step: 230, loss: 0.00822934415191412
/usr/local/lib/python3.7/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=0.9764309764309763, 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: 0.00015048710338305682
step: 10, loss: 0.0001590662432136014
step: 20, loss: 0.00010902166104642674
step: 30, loss: 0.07753544300794601
step: 40, loss: 5.433187470771372e-05
step: 50, loss: 0.006368541624397039
step: 60, loss: 0.033573973923921585
step: 70, loss: 0.00014200303121469915
step: 80, loss: 0.0038134725764393806
step: 90, loss: 0.0003911910462193191
step: 100, loss: 0.0004680190177168697
step: 110, loss: 0.0010791713139042258
step: 120, loss: 0.01900814287364483
step: 130, loss: 0.007808993570506573
step: 140, loss: 0.03417697548866272
step: 150, loss: 0.0016094193561002612
step: 160, loss: 0.0006135560106486082
step: 170, loss: 7.292890222743154e-05
step: 180, loss: 0.0034455035347491503
step: 190, loss: 0.00044962167157791555
step: 200, loss: 0.000540879846084863
step: 210, loss: 0.00011921089026145637
step: 220, loss: 0.0008533028303645551
step: 230, loss: 0.010270604863762856
/usr/local/lib/python3.7/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=0.978865406006674, f1=0.9668874172185431,
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: 0.00020729840616695583
step: 10, loss: 0.0013269613264128566
step: 20, loss: 0.002629365073516965
step: 30, loss: 0.004807574208825827
step: 40, loss: 0.02163321152329445
step: 50, loss: 0.006607250310480595
step: 60, loss: 0.0001716548576951027
step: 70, loss: 0.0022251037880778313
step: 80, loss: 0.00256132404319942
step: 90, loss: 8.823967073112726e-05
step: 100, loss: 0.0016177501529455185
step: 110, loss: 0.0005122284055687487
step: 120, loss: 0.0030243678484112024
step: 130, loss: 4.6973698772490025e-05
step: 140, loss: 9.802790737012401e-05
step: 150, loss: 9.17767538339831e-05
step: 160, loss: 0.00019040561164729297
step: 170, loss: 0.026523977518081665
step: 180, loss: 0.0002646190405357629
step: 190, loss: 8.871655154507607e-05
step: 200, loss: 0.0006456031696870923
step: 210, loss: 0.00025226600700989366
step: 220, loss: 0.0002428351726848632
step: 230, loss: 0.013355166651308537
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9797297297297298, f1=0.9730941704035874,
best_f1=0.9730941704035874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006811738712713122
step: 10, loss: 0.00039252996793948114
step: 20, loss: 6.346527516143396e-05
step: 30, loss: 0.00067418586695556677
step: 40, loss: 9.211315773427486e-05
step: 50, loss: 0.0015791971236467361
step: 60, loss: 0.012508720159530064
step: 70, loss: 6.06961220910307e-05
step: 80, loss: 9.806306479731575e-05
```

```
step: 90, loss: 0.00012345508730504662
step: 100, loss: 6.763023702660576e-05
step: 110, loss: 0.0001863437646534294
step: 120, loss: 0.003998644184321165
step: 130, loss: 7.225629087770358e-05
step: 140, loss: 8.36329345474951e-05
step: 150, loss: 0.00040873128455132246
step: 160, loss: 0.043481480330228806
step: 170, loss: 6.49747671559453e-05
step: 180, loss: 9.791361662792042e-05
step: 190, loss: 0.00018362500122748315
step: 200, loss: 0.015155375935137272
step: 210, loss: 7.28646555216983e-05
step: 220, loss: 0.03733539581298828
step: 230, loss: 0.00011335185263305902
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9819819819819819, 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: 0.0014536096714437008
step: 10, loss: 0.00030195462750270963
step: 20, loss: 6.649147690040991e-05
step: 30, loss: 0.02883448638021946
step: 40, loss: 0.00012483002501539886
step: 50, loss: 6.230505096027628e-05
step: 60, loss: 6.339586252579466e-05
step: 70, loss: 4.2083440348505974e-05
step: 80, loss: 6.159111944725737e-05
step: 90, loss: 6.890692020533606e-05
step: 100, loss: 0.0005073418724350631
step: 110, loss: 0.00030506146140396595
step: 120, loss: 0.00015835431986488402
step: 130, loss: 0.0003400950809009373
step: 140, loss: 0.000446435937192291
step: 150, loss: 0.007011145353317261
step: 160, loss: 6.344803114188835e-05
step: 170, loss: 6.683271931251511e-05
step: 180, loss: 0.0017552273347973824
step: 190, loss: 6.905877671670169e-05
step: 200, loss: 0.00010985824337694794
```

```
step: 210, loss: 6.858664710307494e-05
step: 220, loss: 0.00015248214185703546
step: 230, loss: 9.29843881749548e-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=0.9808773903262092, f1=0.9707865168539327,
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: 5.0904352974612266e-05
step: 10, loss: 3.224904867238365e-05
step: 20, loss: 0.01531339343637228
step: 30, loss: 0.0008259903406724334
step: 40, loss: 0.0001241398131242022
step: 50, loss: 0.0003456832200754434
step: 60, loss: 0.00012648009578697383
step: 70, loss: 0.003330962499603629
step: 80, loss: 6.4557243604213e-05
step: 90, loss: 9.303188562626019e-05
step: 100, loss: 0.02787589468061924
step: 110, loss: 0.00010062653745990247
step: 120, loss: 0.0003613603475969285
step: 130, loss: 6.573714927071705e-05
step: 140, loss: 7.146298594307154e-05
step: 150, loss: 5.614130714093335e-05
step: 160, loss: 5.957496614428237e-05
step: 170, loss: 5.0563947297632694e-05
step: 180, loss: 4.390043977764435e-05
step: 190, loss: 0.0009007397457025945
step: 200, loss: 7.635508518433198e-05
step: 210, loss: 0.013068235479295254
step: 220, loss: 6.681164813926443e-05
step: 230, loss: 3.911684689228423e-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=0.9820224719101124, f1=0.972972972972973,
best_f1=0.972972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.131531750317663e-05
```

```
step: 10, loss: 5.2350122132338583e-05
step: 20, loss: 6.138358730822802e-05
step: 30, loss: 0.00024211584241129458
step: 40, loss: 0.00011174892279086635
step: 50, loss: 4.989783701603301e-05
step: 60, loss: 7.79423862695694e-05
step: 70, loss: 4.251160498824902e-05
step: 80, loss: 0.1316334307193756
step: 90, loss: 2.881081854866352e-05
step: 100, loss: 0.0024032515939325094
step: 110, loss: 6.879808643134311e-05
step: 120, loss: 4.4556105422088876e-05
step: 130, loss: 5.313370274961926e-05
step: 140, loss: 0.03174380958080292
step: 150, loss: 8.935168443713337e-05
step: 160, loss: 5.809970389236696e-05
step: 170, loss: 5.5166448873933405e-05
step: 180, loss: 0.00012395338853821158
step: 190, loss: 0.0035653163213282824
step: 200, loss: 0.00010745965846581385
step: 210, loss: 0.016267335042357445
step: 220, loss: 0.022939663380384445
step: 230, loss: 0.0016124005196616054
/usr/local/lib/python3.7/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=0.9820224719101124, f1=0.9718785151856018,
best_f1=0.972972972972973
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
```

```
loss_scale                : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:08, 281.61it/s]
load_f1 = 0.9755011135857461
real_f1 = 0.9732739420935412
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, 369.81it/s]
```

4.3.2 DBLP-GoogleScholar - Running the matcher

```python
# 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   : 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: 0.7899526953697205
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4582822918891907

```
step: 20, loss: 0.5097065567970276
step: 30, loss: 0.43258506059646606
step: 40, loss: 0.32118362188339233
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.26988837122917175
step: 60, loss: 0.16660654544830322
step: 70, loss: 0.2875852882862091
step: 80, loss: 0.1217379942536354
step: 90, loss: 0.07188457250595093
step: 100, loss: 0.3403911888599396
step: 110, loss: 0.035923514515161514
step: 120, loss: 0.048886679112911224
step: 130, loss: 0.05522674322128296
step: 140, loss: 0.16464798152446747
step: 150, loss: 0.023787008598446846
step: 160, loss: 0.23431681096553802
step: 170, loss: 0.19470429420471191
step: 180, loss: 0.12852852046489716
step: 190, loss: 0.06272514909505844
step: 200, loss: 0.10945910960435867
step: 210, loss: 0.04277154803276062
step: 220, loss: 0.13372738659381866
step: 230, loss: 0.12209965288639069
step: 240, loss: 0.08900922536849976
step: 250, loss: 0.0842413604259491
step: 260, loss: 0.029711971059441566
step: 270, loss: 0.016037601977586746
step: 280, loss: 0.09668541699647903
step: 290, loss: 0.10688420385122299
step: 300, loss: 0.16231006383895874
step: 310, loss: 0.06622314453125
step: 320, loss: 0.03254642337560654
step: 330, loss: 0.08508697152137756
step: 340, loss: 0.1867525279521942
step: 350, loss: 0.1054045632481575
step: 360, loss: 0.032721247524023056
step: 370, loss: 0.14652541279792786
step: 380, loss: 0.1654195785522461
step: 390, loss: 0.01171058602631092
step: 400, loss: 0.008796203881502151
step: 410, loss: 0.01834571175277233
step: 420, loss: 0.022292139008641243
step: 430, loss: 0.16066385805606842
step: 440, loss: 0.055754851549863815
step: 450, loss: 0.03426002711057663
step: 460, loss: 0.15950535237789154
step: 470, loss: 0.28255075216293335
step: 480, loss: 0.41377636790275574
```

```
step: 490, loss: 0.03783801570534706
step: 500, loss: 0.007689372170716524
step: 510, loss: 0.04629778489470482
step: 520, loss: 0.03329193964600563
step: 530, loss: 0.09409263730049133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9325210871602625, f1=0.9324577861163227,
best_f1=0.9324577861163227
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13293246924877167
step: 10, loss: 0.20997050404548645
step: 20, loss: 0.12012692540884018
step: 30, loss: 0.08696844428777695
step: 40, loss: 0.008418618701398373
step: 50, loss: 0.10962731391191483
step: 60, loss: 0.10879073292016983
step: 70, loss: 0.2080247402191162
step: 80, loss: 0.042865537106990814
step: 90, loss: 0.025168081745505333
step: 100, loss: 0.3421240746974945
step: 110, loss: 0.04753530025482178
step: 120, loss: 0.04880908504128456
step: 130, loss: 0.011924207210540771
step: 140, loss: 0.017828527837991714
step: 150, loss: 0.1500030905008316
step: 160, loss: 0.02255896106362343
step: 170, loss: 0.027109960094094276
step: 180, loss: 0.01616443693637848
step: 190, loss: 0.01612108014523983
step: 200, loss: 0.043669842183589935
step: 210, loss: 0.021640364080667496
step: 220, loss: 0.08274465799331665
step: 230, loss: 0.014001058414578438
step: 240, loss: 0.1693524867296219
step: 250, loss: 0.06693047285079956
step: 260, loss: 0.009020951576530933
step: 270, loss: 0.4210754930973053
step: 280, loss: 0.09749126434326172
step: 290, loss: 0.10324709862470627
step: 300, loss: 0.024885879829525948
```

```
step: 310, loss: 0.04661991074681282
step: 320, loss: 0.12200778722763062
step: 330, loss: 0.04775971546769142
step: 340, loss: 0.021676212549209595
step: 350, loss: 0.048599254339933395
step: 360, loss: 0.0096882693469524838
step: 370, loss: 0.011207865551114082
step: 380, loss: 0.05877208709716797
step: 390, loss: 0.022154200822114944
step: 400, loss: 0.0430438369512558
step: 410, loss: 0.007336434908211231
step: 420, loss: 0.07425142824649811
step: 430, loss: 0.022879716008901596
step: 440, loss: 0.01072340365499258
step: 450, loss: 0.08035391569137573
step: 460, loss: 0.2486741542816162
step: 470, loss: 0.06603129953145981
step: 480, loss: 0.21774335205554962
step: 490, loss: 0.04083145782351494
step: 500, loss: 0.049071215093135834
step: 510, loss: 0.07438022643327713
step: 520, loss: 0.05358906090259552
step: 530, loss: 0.1422453671693802
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9450249886415266, f1=0.9340009103322713,
best_f1=0.9340009103322713
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009529536589980125
step: 10, loss: 0.09103486686944962
step: 20, loss: 0.16394037008285522
step: 30, loss: 0.2973191440105438
step: 40, loss: 0.08466677367687225
step: 50, loss: 0.04144713282585144
step: 60, loss: 0.042200099002122879
step: 70, loss: 0.15658272802829742
step: 80, loss: 0.16125182807445526
step: 90, loss: 0.12828078866004944
step: 100, loss: 0.03706570342183113
step: 110, loss: 0.07433374226093292
step: 120, loss: 0.023808415979146957
```

```
step: 130, loss: 0.01124855037778616
step: 140, loss: 0.03051701746881008
step: 150, loss: 0.021737152710556984
step: 160, loss: 0.019890615716576576
step: 170, loss: 0.007233778014779091
step: 180, loss: 0.0246674045920372
step: 190, loss: 0.013464929535984993
step: 200, loss: 0.03251214697957039
step: 210, loss: 0.007176556158810854
step: 220, loss: 0.013406707905232906
step: 230, loss: 0.05687447637319565
step: 240, loss: 0.025787638500332832
step: 250, loss: 0.006509039551019669
step: 260, loss: 0.10094044357538223
step: 270, loss: 0.002771698869764805
step: 280, loss: 0.005954332184046507
step: 290, loss: 0.06752683967351913
step: 300, loss: 0.08968880772590637
step: 310, loss: 0.12474768608808517
step: 320, loss: 0.1800357699394226
step: 330, loss: 0.02273782715201378
step: 340, loss: 0.00201878952793777
step: 350, loss: 0.028507528826594353
step: 360, loss: 0.005830579437315464
step: 370, loss: 0.004892466124147177
step: 380, loss: 0.06276097893714905
step: 390, loss: 0.039436206221580505
step: 400, loss: 0.018439704552292824
step: 410, loss: 0.01865256577730179
step: 420, loss: 0.04552091658115387
step: 430, loss: 0.019330715760588646
step: 440, loss: 0.08133628964424133
step: 450, loss: 0.013233962468802929
step: 460, loss: 0.15495501458644867
step: 470, loss: 0.008943921886384487
step: 480, loss: 0.02060159668326378
step: 490, loss: 0.05265270173549652
step: 500, loss: 0.13876189291477203
step: 510, loss: 0.0026069015730172396
step: 520, loss: 0.006623651832342148
step: 530, loss: 0.05784660577774048
/usr/local/lib/python3.7/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=0.9392675011590171, f1=0.9369786839666358,
best_f1=0.9340009103322713
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027115163393318653
step: 10, loss: 0.09275930374860764
step: 20, loss: 0.05830986797809601
step: 30, loss: 0.0025543447118252516
step: 40, loss: 0.003077197354286909
step: 50, loss: 0.008641431108117104
step: 60, loss: 0.0005405591218732297
step: 70, loss: 0.002897093305364251
step: 80, loss: 0.03123973309993744
step: 90, loss: 0.04843307286500931
step: 100, loss: 0.003929590806365013
step: 110, loss: 0.003361461451277137
step: 120, loss: 0.016886139288544655
step: 130, loss: 0.016181115061044693
step: 140, loss: 0.11194345355033875
step: 150, loss: 0.008162891492247581
step: 160, loss: 0.002293840516358614
step: 170, loss: 0.03130580484867096
step: 180, loss: 0.013312377966940403
step: 190, loss: 0.08149241656064987
step: 200, loss: 0.03666439279913902
step: 210, loss: 0.12884746491909027
step: 220, loss: 0.012121972627937794
step: 230, loss: 0.1793648600578308
step: 240, loss: 0.005096865817904472
step: 250, loss: 0.0026308197993785143
step: 260, loss: 0.10459224134683609
step: 270, loss: 0.008534757420420647
step: 280, loss: 0.008167284540832043
step: 290, loss: 0.0058370172046124935
step: 300, loss: 0.0011296941665932536
step: 310, loss: 0.02605828084051609
step: 320, loss: 0.0919630154967308
step: 330, loss: 0.09409300982952118
step: 340, loss: 0.018743157386779785
step: 350, loss: 0.0045569404028356075
step: 360, loss: 0.0021643205545842648
step: 370, loss: 0.049077026546001434
step: 380, loss: 0.019282551482319832
step: 390, loss: 0.21203212440013885
step: 400, loss: 0.007540116552263498
step: 410, loss: 0.035529885441064835
step: 420, loss: 0.07295354455709457
step: 430, loss: 0.018316539004445076
step: 440, loss: 0.12113755941390991
step: 450, loss: 0.01723867654800415
step: 460, loss: 0.002509991405531764
```

```
step: 470, loss: 0.09414222091436386
step: 480, loss: 0.014366932213306427
step: 490, loss: 0.05669711157679558
step: 500, loss: 0.01929553784430027
step: 510, loss: 0.1033712700009346
step: 520, loss: 0.003091109450906515
step: 530, loss: 0.004782940726727247
/usr/local/lib/python3.7/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=0.9410664172123478, f1=0.9364485981308411,
best_f1=0.9340009103322713
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003039844101294875
step: 10, loss: 0.02086598053574562
step: 20, loss: 0.018778890371322632
step: 30, loss: 0.0022077003959566355
step: 40, loss: 0.06515401601791382
step: 50, loss: 0.0048042829148471355
step: 60, loss: 0.002279737265780568
step: 70, loss: 0.0066169314086437225
step: 80, loss: 0.007613468915224075
step: 90, loss: 0.02146734483540058
step: 100, loss: 0.0005261427722871304
step: 110, loss: 0.01810154691338539
step: 120, loss: 0.00442897342145443
step: 130, loss: 0.0034481592010706663
step: 140, loss: 0.06912118196487427
step: 150, loss: 0.08698750287294388
step: 160, loss: 0.16775505244731903
step: 170, loss: 0.06019534915685654
step: 180, loss: 0.0009971046820282936
step: 190, loss: 0.0034215073101222515
step: 200, loss: 0.12635713815689087
step: 210, loss: 0.01679551973938942
step: 220, loss: 0.0009075651178136468
step: 230, loss: 0.0026403216179460287
step: 240, loss: 0.006970325950533152
step: 250, loss: 0.0026564286090433598
step: 260, loss: 0.06428586691617966
step: 270, loss: 0.004178880713880062
step: 280, loss: 0.1569664478302002
step: 290, loss: 0.003036474110558629
step: 300, loss: 0.007307430263608694
step: 310, loss: 0.017457427456974983
step: 320, loss: 0.007114786189049482
```

```
step: 330, loss: 0.0019634542986750603
step: 340, loss: 0.02218116819858551
step: 350, loss: 0.016797736287117004
step: 360, loss: 0.0050171734765172005
step: 370, loss: 0.002246126066893339
step: 380, loss: 0.03407984972000122
step: 390, loss: 0.009404842741787434
step: 400, loss: 0.0775105208158493
step: 410, loss: 0.010598755441606045
step: 420, loss: 0.003966172691434622
step: 430, loss: 0.008573190309107304
step: 440, loss: 0.0965854674577713
step: 450, loss: 0.015200180932879448
step: 460, loss: 0.005983081180602312
step: 470, loss: 0.024719256907701492
step: 480, loss: 0.06013109162449837
step: 490, loss: 0.01436984445899725
step: 500, loss: 0.045727021992206573
step: 510, loss: 0.02448856644332409
step: 520, loss: 0.0032190021593123674
step: 530, loss: 0.005332747474312782
/usr/local/lib/python3.7/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=0.9387186629526463, f1=0.9308118081180812,
best_f1=0.9340009103322713
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0050053502507507801
step: 10, loss: 0.0020299770403653383
step: 20, loss: 0.00045673068962059915
step: 30, loss: 0.005103544797748327
step: 40, loss: 0.004570073448121548
step: 50, loss: 0.012347813695669174
step: 60, loss: 0.0010093295713886619
step: 70, loss: 0.02928268536925316
step: 80, loss: 0.001220058649778366
step: 90, loss: 0.0018764968262985349
step: 100, loss: 0.04862991347908974
step: 110, loss: 0.021063974127173424
step: 120, loss: 0.002643525367602706
step: 130, loss: 0.08760403096675873
step: 140, loss: 0.00288395956158638
step: 150, loss: 0.006136584095656872
step: 160, loss: 0.007256970275193453
step: 170, loss: 0.09349729865789413
step: 180, loss: 0.00011712258856277913
```

```
step: 190, loss: 0.015462665818631649
step: 200, loss: 0.0008390233269892633
step: 210, loss: 0.0009103566990233958
step: 220, loss: 0.06096995621919632
step: 230, loss: 0.00024325006233993918
step: 240, loss: 0.00028377119451761246
step: 250, loss: 0.0019657539669424295
step: 260, loss: 0.01713928021490574
step: 270, loss: 0.00011958600953221321
step: 280, loss: 0.001075276522897184
step: 290, loss: 0.0003695294726639986
step: 300, loss: 0.001794385607354343
step: 310, loss: 0.001425692462362349
step: 320, loss: 0.10367248207330704
step: 330, loss: 0.0014459906378760934
step: 340, loss: 0.0011777670588344336
step: 350, loss: 0.1331910640001297
step: 360, loss: 0.00012758791854139417
step: 370, loss: 0.0036911056376993656
step: 380, loss: 0.024805035442113876
step: 390, loss: 0.009235099889338017
step: 400, loss: 0.00027406559092924
step: 410, loss: 0.0003184730012435466
step: 420, loss: 0.01973697729408741
step: 430, loss: 0.0002665214997250587
step: 440, loss: 0.015605033375322819
step: 450, loss: 0.012098448351025581
step: 460, loss: 0.005331987980753183
step: 470, loss: 0.06239913031458855
step: 480, loss: 0.004628350958228111
step: 490, loss: 0.001704445923678577
step: 500, loss: 0.0033186653163284063
step: 510, loss: 6.969830428715795e-05
step: 520, loss: 0.003525621723383665
step: 530, loss: 0.009702039882540703
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9455719557195572, f1=0.9355586462679647,
best_f1=0.9355586462679647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004484924487769604
```

```
step: 10, loss: 0.03163626417517662
step: 20, loss: 0.08518822491168976
step: 30, loss: 0.00591913890093565
step: 40, loss: 0.0018127188086509705
step: 50, loss: 0.009909965097904205
step: 60, loss: 0.019982846453785896
step: 70, loss: 0.020938608795404434
step: 80, loss: 0.01701195351779461
step: 90, loss: 0.00013702051364816725
step: 100, loss: 0.0007739478605799377
step: 110, loss: 0.1555033177137375
step: 120, loss: 5.052053165854886e-05
step: 130, loss: 0.01461127307265997
step: 140, loss: 0.01245365384966135
step: 150, loss: 0.002377081662416458
step: 160, loss: 0.003824642626568675
step: 170, loss: 0.00547407940030098
step: 180, loss: 0.023729924112558365
step: 190, loss: 0.0004970930167473853
step: 200, loss: 0.004832103382796049
step: 210, loss: 0.0007281530415639281
step: 220, loss: 0.07312382012605667
step: 230, loss: 0.003113019745796919
step: 240, loss: 0.005197568330913782
step: 250, loss: 0.004365443717688322
step: 260, loss: 0.010360793210566044
step: 270, loss: 0.00022276703384704888
step: 280, loss: 0.059677090495824814
step: 290, loss: 0.0027119889855384827
step: 300, loss: 0.0004043072694912553
step: 310, loss: 0.00034370197681710124
step: 320, loss: 0.026241270825266838
step: 330, loss: 0.0027011169586330652
step: 340, loss: 0.006395857781171799
step: 350, loss: 0.00012057242565788329
step: 360, loss: 0.005632249172769375
step: 370, loss: 0.0017445194534957409
step: 380, loss: 0.0034848216455429792
step: 390, loss: 0.004604901187121868
step: 400, loss: 0.00017629079229664057
step: 410, loss: 0.0013620859244838357
step: 420, loss: 0.000544318463653326
step: 430, loss: 0.00010356913844589144
step: 440, loss: 0.04711730778217316
step: 450, loss: 0.002423232886940241
step: 460, loss: 0.0047420053742825985
step: 470, loss: 0.1100202202796936
step: 480, loss: 0.002830084413290024
```

```
step: 490, loss: 0.008904820308089256
step: 500, loss: 0.0001992639881791547
step: 510, loss: 0.0018730416195467114
step: 520, loss: 0.04974367469549179
step: 530, loss: 0.0015120763564482331
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9488243430152145, f1=0.9394495412844037,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007482678163796663
step: 10, loss: 0.03444268926978111
step: 20, loss: 0.008849021978676319
step: 30, loss: 0.021304728463292122
step: 40, loss: 0.0005402773967944086
step: 50, loss: 0.0011531885247677565
step: 60, loss: 0.0017048283480107784
step: 70, loss: 0.00806475430727005
step: 80, loss: 0.0004067239642608911
step: 90, loss: 0.00011729869584087282
step: 100, loss: 0.0013289295602589846
step: 110, loss: 0.000731108128093183
step: 120, loss: 0.00021256563195493072
step: 130, loss: 0.00043121696216985583
step: 140, loss: 4.827453085454181e-05
step: 150, loss: 0.0007888741674833
step: 160, loss: 0.0013968293787911534
step: 170, loss: 0.0006952413823455572
step: 180, loss: 0.005000223405659199
step: 190, loss: 0.0031509019900113344
step: 200, loss: 0.0008038745145313442
step: 210, loss: 0.0031395810656249523
step: 220, loss: 0.008311126381158829
step: 230, loss: 0.0038515778724104166
step: 240, loss: 0.01438638661056757
step: 250, loss: 0.0016265091253444552
step: 260, loss: 0.0008308027754537761
step: 270, loss: 0.00019126635743305087
step: 280, loss: 0.0005439367960207164
step: 290, loss: 0.00038811436388641596
step: 300, loss: 4.6462355385301635e-05
```

```
step: 310, loss: 0.02250364050269127
step: 320, loss: 0.0002676319272723049
step: 330, loss: 0.03824685141444206
step: 340, loss: 0.002214878099039197
step: 350, loss: 0.002951979171484709
step: 360, loss: 0.002517087385058403
step: 370, loss: 0.003964274190366268
step: 380, loss: 0.011923235841095448
step: 390, loss: 9.395962115377188e-05
step: 400, loss: 0.03986869752407074
step: 410, loss: 0.0010939823696389794
step: 420, loss: 0.00020325841614976525
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 430, loss: 0.010960433632135391
step: 440, loss: 0.0006065089255571365
step: 450, loss: 0.0022308651823550463
step: 460, loss: 0.005809019319713116
step: 470, loss: 0.00019939335729842633
step: 480, loss: 0.004455437883734703
step: 490, loss: 0.00033261749194934964
step: 500, loss: 0.004776053596287966
step: 510, loss: 0.00016489367408212274
step: 520, loss: 0.0005832537426613271
step: 530, loss: 0.0003957879962399602
/usr/local/lib/python3.7/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=0.9434137291280148, f1=0.9309225776541493,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007826478104107082
step: 10, loss: 0.015860671177506447
step: 20, loss: 0.0008983499137684703
step: 30, loss: 0.005445664282888174
step: 40, loss: 0.0015953583642840385
step: 50, loss: 0.0011531597701832652
step: 60, loss: 0.0012895376421511173
step: 70, loss: 0.11043717712163925
step: 80, loss: 0.0003677415370475501
step: 90, loss: 0.006268879398703575
step: 100, loss: 0.00026353204157203436
step: 110, loss: 0.0017281000036746264
step: 120, loss: 4.603140041581355e-05
step: 130, loss: 0.001964018912985921
step: 140, loss: 0.019176891073584557
step: 150, loss: 0.0004793773696292192
```

```
step: 160, loss: 0.004792413674294949
step: 170, loss: 0.0005015115602873266
step: 180, loss: 0.002350538270547986
step: 190, loss: 0.11829308420419693
step: 200, loss: 0.03302778676152229
step: 210, loss: 0.01089375838637352
step: 220, loss: 0.11862008273601532
step: 230, loss: 0.0013191265752539039
step: 240, loss: 0.006511914078146219
step: 250, loss: 0.0029925834387540817
step: 260, loss: 0.0019940577913075686
step: 270, loss: 0.020312169566750526
step: 280, loss: 0.00421216432005167
step: 290, loss: 0.0006619667983613908
step: 300, loss: 0.0016288962215185165
step: 310, loss: 0.00014072266640141606
step: 320, loss: 0.0012726141139864922
step: 330, loss: 0.021421775221824646
step: 340, loss: 0.0002061589912045747
step: 350, loss: 0.00018013401131611317
step: 360, loss: 0.06308867037296295
step: 370, loss: 0.00040437112329527736
step: 380, loss: 0.0006134265568107367
step: 390, loss: 0.0004906300455331802
step: 400, loss: 0.020032323896884918
step: 410, loss: 0.0001340792077826336
step: 420, loss: 0.0008655228884890676
step: 430, loss: 4.2693642171798274e-05
step: 440, loss: 0.0003606590034905821
step: 450, loss: 0.00013014981232117862
step: 460, loss: 0.0001761910825734958
step: 470, loss: 3.2792348065413535e-05
step: 480, loss: 0.00012477021664381027
step: 490, loss: 8.569294004701078e-05
step: 500, loss: 0.037038885056972504
step: 510, loss: 0.0004473023291211575
step: 520, loss: 6.934957491466776e-05
step: 530, loss: 0.00013591666356660426
/usr/local/lib/python3.7/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=0.9456721915285451, f1=0.9368709972552608,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006129913032054901
step: 10, loss: 0.00012485688785091043
```

```
step: 20, loss: 0.0003960731264669448
step: 30, loss: 0.004679666832089424
step: 40, loss: 0.0047174436040222645
step: 50, loss: 0.00035101911635138094
step: 60, loss: 0.0005196290439926088
step: 70, loss: 0.0012670308351516724
step: 80, loss: 0.0008456241339445114
step: 90, loss: 0.010521183721721172
step: 100, loss: 4.268510747351684e-05
step: 110, loss: 0.0006911368109285831
step: 120, loss: 0.00017420756921637803
step: 130, loss: 7.768847717670724e-05
step: 140, loss: 0.00020436795603018254
step: 150, loss: 6.653823220403865e-05
step: 160, loss: 0.0019280043197795749
step: 170, loss: 0.00022802213788963854
step: 180, loss: 0.03010343760251999
step: 190, loss: 0.00012920999142806977
step: 200, loss: 8.308464020956308e-05
step: 210, loss: 5.4213087423704565e-05
step: 220, loss: 0.00038038057391531765
step: 230, loss: 0.00029433725285343826
step: 240, loss: 3.393942097318359e-05
step: 250, loss: 0.00015018138219602406
step: 260, loss: 0.007953476160764694
step: 270, loss: 4.787874422618188e-05
step: 280, loss: 0.0005928560276515782
step: 290, loss: 0.00021471965010277927
step: 300, loss: 0.0009229819988831878
step: 310, loss: 0.0046999105252325535
step: 320, loss: 0.007192731369286776
step: 330, loss: 6.657276389887556e-05
step: 340, loss: 0.05031384900212288
step: 350, loss: 8.342790533788502e-05
step: 360, loss: 0.008480604737997055
step: 370, loss: 0.0010569622972980142
step: 380, loss: 0.001189161790534854
step: 390, loss: 5.1678638556040823e-05
step: 400, loss: 0.00015798138338141143
step: 410, loss: 0.0008827283163554966
step: 420, loss: 0.005494699347764254
step: 430, loss: 3.7615103792632e-05
step: 440, loss: 3.5406181268626824e-05
step: 450, loss: 0.032467618584632874
step: 460, loss: 0.002148744184523821
step: 470, loss: 0.00010561854287516326
step: 480, loss: 0.0038629709742963314
step: 490, loss: 0.0758393406867981
```

```
step: 500, loss: 0.014443120919167995
step: 510, loss: 0.00043165188981220126
step: 520, loss: 0.015571050345897675
step: 530, loss: 0.00015309751324821264
/usr/local/lib/python3.7/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=0.9415614773258532, f1=0.9328984156570364,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0076079657301306725
step: 10, loss: 0.0009907351341098547
step: 20, loss: 5.4088308388600126e-05
step: 30, loss: 0.001994568854570389
step: 40, loss: 0.003620320465415716
step: 50, loss: 0.001160894986242056
step: 60, loss: 3.577870666049421e-05
step: 70, loss: 0.00011597530829021707
step: 80, loss: 3.7532947317231447e-05
step: 90, loss: 0.0008437251672148705
step: 100, loss: 0.011493168771266937
step: 110, loss: 0.0007163500413298607
step: 120, loss: 0.021421590819954872
step: 130, loss: 0.006623863242566585
step: 140, loss: 6.906646740389988e-05
step: 150, loss: 0.0013043336803093553
step: 160, loss: 0.0024158016312867403
step: 170, loss: 0.0025943678338080645
step: 180, loss: 3.006206316058524e-05
step: 190, loss: 0.001938651199452579
step: 200, loss: 0.020233644172549248
step: 210, loss: 0.00022114843886811286
step: 220, loss: 7.10969470674172e-05
step: 230, loss: 0.0003694953629747033
step: 240, loss: 0.0005941505078226328
step: 250, loss: 0.0013224580325186253
step: 260, loss: 5.282769780023955e-05
step: 270, loss: 0.000979081029072404
step: 280, loss: 0.02251484990119934
step: 290, loss: 0.02597075141966343
step: 300, loss: 0.0019828907679766417
step: 310, loss: 7.150899909902364e-05
step: 320, loss: 0.0072050741873681545
step: 330, loss: 0.00400543212890625
step: 340, loss: 0.0004218074318487197
step: 350, loss: 0.0032925123814493418
```

```
step: 360, loss: 0.00014981265121605247
step: 370, loss: 2.4377408408327028e-05
step: 380, loss: 3.715306593221612e-05
step: 390, loss: 0.00047376833390444517
step: 400, loss: 5.176264312467538e-05
step: 410, loss: 8.953808719525114e-05
step: 420, loss: 0.0005130111239850521
step: 430, loss: 0.00014748259854968637
step: 440, loss: 3.9975082472665235e-05
step: 450, loss: 0.0001879826741060242
step: 460, loss: 0.0017306070076301694
step: 470, loss: 0.0003543018829077482
step: 480, loss: 0.00014010605809744447
step: 490, loss: 8.526545570930466e-05
step: 500, loss: 6.370549090206623e-05
step: 510, loss: 0.00048359876382164657
step: 520, loss: 2.3044227418722585e-05
step: 530, loss: 7.861694029998034e-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=0.9470725995316158, f1=0.9376744186046512,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00027798491646535695
step: 10, loss: 3.5413460864219815e-05
step: 20, loss: 1.513199367764173e-05
step: 30, loss: 0.0016441718908026814
step: 40, loss: 0.0005447995499707758
step: 50, loss: 0.0023041798267513514
step: 60, loss: 2.2038388124201447e-05
step: 70, loss: 0.0007883411017246544
step: 80, loss: 0.0014281650073826313
step: 90, loss: 0.000882828317116946
step: 100, loss: 0.003644702723249793
step: 110, loss: 0.0850265771150589
step: 120, loss: 0.00026456857449375093
step: 130, loss: 4.226506644045003e-05
step: 140, loss: 0.00018846562306862324
step: 150, loss: 0.02120205946266651
step: 160, loss: 0.0004297170671634376
step: 170, loss: 0.0002631053503137082
step: 180, loss: 0.006965699605643749
step: 190, loss: 1.4245399142964743e-05
step: 200, loss: 1.9344932297826745e-05
step: 210, loss: 0.0024037379771471024
```

```
step: 220, loss: 3.515406933729537e-05
step: 230, loss: 0.001576202572323382
step: 240, loss: 0.0003610706771723926
step: 250, loss: 9.720971866045147e-05
step: 260, loss: 0.0019679907709360123
step: 270, loss: 0.0056651984341442585
step: 280, loss: 2.026090987783391e-05
step: 290, loss: 0.008500429801642895
step: 300, loss: 0.01049424335360527
step: 310, loss: 0.00031096884049475193
step: 320, loss: 0.022440945729613304
step: 330, loss: 0.00037950483965687454
step: 340, loss: 0.026816461235284805
step: 350, loss: 0.0037419390864670277
step: 360, loss: 0.012067737989127636
step: 370, loss: 1.9367495042388327e-05
step: 380, loss: 0.0005350707797333598
step: 390, loss: 0.000361282960511744
step: 400, loss: 6.225884862942621e-05
step: 410, loss: 0.016805393621325493
step: 420, loss: 0.0010694051161408424
step: 430, loss: 0.0008085417794063687
step: 440, loss: 0.0003505549975670874
step: 450, loss: 0.009807172231376171
step: 460, loss: 0.002214777749031782
step: 470, loss: 0.0001708879426587373
step: 480, loss: 0.0001310984225710854
step: 490, loss: 0.0012577895540744066
step: 500, loss: 0.002199075184762478
step: 510, loss: 0.001655804575420916
step: 520, loss: 0.008302442729473114
step: 530, loss: 0.000489225029014051
/usr/local/lib/python3.7/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=0.9477124183006537, f1=0.9399161620866325,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000321295257890597
step: 10, loss: 0.00026164125301875174
step: 20, loss: 0.05464356020092964
step: 30, loss: 0.0004172587941866368
step: 40, loss: 0.0003359775000717491
step: 50, loss: 0.0006913218530826271
step: 60, loss: 0.0013199775712564588
step: 70, loss: 0.000369959365343675
```

```
step: 80, loss: 3.3072388760047033e-05
step: 90, loss: 0.0016828137449920177
step: 100, loss: 2.7385622161091305e-05
step: 110, loss: 0.0003894343099091202
step: 120, loss: 0.00045668688835576177
step: 130, loss: 0.0029461286030709743
step: 140, loss: 2.9454455216182396e-05
step: 150, loss: 0.00017579014820512384
step: 160, loss: 0.011829011142253876
step: 170, loss: 0.006133846938610077
step: 180, loss: 0.00042950778151862323
step: 190, loss: 0.00012400899140629917
step: 200, loss: 0.002191167324781418
step: 210, loss: 0.0014031118480488658
step: 220, loss: 0.0015239011263474822
step: 230, loss: 0.028748825192451477
step: 240, loss: 3.6032895877724513e-05
step: 250, loss: 0.02853713557124138
step: 260, loss: 0.00013063970254734159
step: 270, loss: 0.001127908006310463
step: 280, loss: 0.000639180070720613
step: 290, loss: 0.004618664272129536
step: 300, loss: 0.0007568090222775936
step: 310, loss: 0.0037533065769821405
step: 320, loss: 2.7415488148108125e-05
step: 330, loss: 0.0029710386879742146
step: 340, loss: 2.8582324375747703e-05
step: 350, loss: 0.01781616173684597
step: 360, loss: 4.28169769293163e-05
step: 370, loss: 5.421060632215813e-05
step: 380, loss: 0.0026511431206017733
step: 390, loss: 0.0018367258599027991
step: 400, loss: 6.97318246238865e-05
step: 410, loss: 0.00028036610456183553
step: 420, loss: 0.00039602292235940695
step: 430, loss: 0.001595427980646491
step: 440, loss: 0.0004685370367951691
step: 450, loss: 0.0008687087101861835
step: 460, loss: 0.0006376600358635187
step: 470, loss: 0.02933507226407528
step: 480, loss: 0.0009277024655602872
step: 490, loss: 0.0003859552671201527
step: 500, loss: 0.0006906857015565038
step: 510, loss: 2.231725193269085e-05
step: 520, loss: 6.478362047346309e-05
step: 530, loss: 7.468117109965533e-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=0.9465290806754221, f1=0.9363295880149813,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.861781755811535e-05
step: 10, loss: 5.8332101616542786e-05
step: 20, loss: 4.192719279672019e-05
step: 30, loss: 0.0004749791114591062
step: 40, loss: 0.0005755347083322704
step: 50, loss: 0.0027721519581973553
step: 60, loss: 1.8372800695942715e-05
step: 70, loss: 5.6357679568463936e-05
step: 80, loss: 0.002042583655565977
step: 90, loss: 7.109731814125553e-05
step: 100, loss: 0.009211871773004532
step: 110, loss: 8.46499387989752e-05
step: 120, loss: 0.00222774944268167
step: 130, loss: 0.00023745255020912737
step: 140, loss: 0.013518552295863628
step: 150, loss: 0.00021418971300590783
step: 160, loss: 0.001512877526693046
step: 170, loss: 0.00020859380310866982
step: 180, loss: 0.0026219955179840326
step: 190, loss: 0.00016370337107218802
step: 200, loss: 0.000340828497428447
step: 210, loss: 4.740913936984725e-05
step: 220, loss: 9.406059689354151e-05
step: 230, loss: 0.0009815727826207876
step: 240, loss: 0.0004446574312169105
step: 250, loss: 0.008708293549716473
step: 260, loss: 3.857758565573022e-05
step: 270, loss: 0.003393663791939616
step: 280, loss: 6.203652446856722e-05
step: 290, loss: 6.146440136944875e-05
step: 300, loss: 0.0033232253044843674
step: 310, loss: 7.598701631650329e-05
step: 320, loss: 0.000396952498704195
step: 330, loss: 0.002225564094260335
step: 340, loss: 3.190291317878291e-05
step: 350, loss: 0.0006853786180727184
step: 360, loss: 0.002013824414461851
step: 370, loss: 0.00012129824608564377
step: 380, loss: 0.0004956990596838295
step: 390, loss: 4.571394310914911e-05
step: 400, loss: 2.3661772502237e-05
step: 410, loss: 0.00010766838386189193
```

```
step: 420, loss: 0.0003578835166990757
step: 430, loss: 0.0011710120597854257
step: 440, loss: 2.0712132027256303e-05
step: 450, loss: 3.1265135476132855e-05
step: 460, loss: 0.00027786404825747013
step: 470, loss: 2.8631018722080626e-05
step: 480, loss: 5.335634705261327e-05
step: 490, loss: 0.00016490279813297093
step: 500, loss: 0.00012111567048123106
step: 510, loss: 0.0006106672226451337
step: 520, loss: 3.976172229158692e-05
step: 530, loss: 0.0004794921260327101
/usr/local/lib/python3.7/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=0.9441860465116279, f1=0.937528921795465,
best_f1=0.9394495412844037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011950857151532546
step: 10, loss: 2.480171860952396e-05
step: 20, loss: 0.0010215624934062362
step: 30, loss: 0.0014903830597177148
step: 40, loss: 2.0592326109181158e-05
step: 50, loss: 0.0012784532736986876
step: 60, loss: 1.4066570656723343e-05
step: 70, loss: 0.00034515580045990646
step: 80, loss: 1.6178686564671807e-05
step: 90, loss: 1.035997411236167e-05
step: 100, loss: 1.745636109262705e-05
step: 110, loss: 0.008312291465699673
step: 120, loss: 0.0013526123948395252
step: 130, loss: 8.161050209309906e-05
step: 140, loss: 1.6070654965005815e-05
step: 150, loss: 0.00018392610945738852
step: 160, loss: 6.504428165499121e-05
step: 170, loss: 4.0291437471751124e-05
step: 180, loss: 8.597902342444286e-05
step: 190, loss: 0.0005582306766882539
step: 200, loss: 3.830537025351077e-05
step: 210, loss: 4.145013735978864e-05
step: 220, loss: 9.370478801429272e-05
step: 230, loss: 8.983108273241669e-05
step: 240, loss: 0.001462809625081718
step: 250, loss: 0.00023056127247400582
step: 260, loss: 0.00208052690140903
step: 270, loss: 7.668169564567506e-05
```

```
step: 280, loss: 0.0010132086463272572
step: 290, loss: 0.0011041470570489764
step: 300, loss: 0.0006779562681913376
step: 310, loss: 0.00016577520000282675
step: 320, loss: 3.862397716147825e-05
step: 330, loss: 7.854526484152302e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 340, loss: 0.00016364105977118015
step: 350, loss: 3.010864747921005e-05
step: 360, loss: 4.8841306124813855e-05
step: 370, loss: 2.957524520752486e-05
step: 380, loss: 6.522875628434122e-05
step: 390, loss: 0.0017415526090189815
step: 400, loss: 0.00043007038766518235
step: 410, loss: 0.0014892799081280828
step: 420, loss: 3.1025887437863275e-05
step: 430, loss: 1.8454689779900946e-05
step: 440, loss: 0.004408976528793573
step: 450, loss: 0.00163573632016778
step: 460, loss: 0.0001272850640816614
step: 470, loss: 0.00019584815890993923
step: 480, loss: 7.02920151525177e-05
step: 490, loss: 0.042314086109399796
step: 500, loss: 0.0010940697975456715
step: 510, loss: 0.008213063701987267
step: 520, loss: 0.00361249758861959
step: 530, loss: 1.9203249394195154e-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=0.9446768944676894, f1=0.9376443418013857,
best_f1=0.9394495412844037
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)…
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
```

```
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale            : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:13, 422.93it/s]
load_f1 = 0.9482999534233814
real_f1 = 0.9486940298507462
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:10, 430.56it/s]
```