pytorch implementation of video captioning

recommend installing pytorch and python packages using Anaconda

requirements

- cuda
- pytorch 0.4.0
- python3
- ffmpeg (can install using anaconda)

python packages

- tqdm
- pillow
- pretrainedmodels
- nltk

Data

MSR-VTT. Test video doesn't have captions, so I spilit train-viedo to train/val/test. Extract and put them in `./data/` directory

- train-video: download link
- test-video: download link
- json info of train-video: download link
- json info of test-video: download link

Options

all default options are defined in opt.py or corresponding code file, change them for your like.

Acknowledgements

Some code refers to ImageCaptioning.pytorch

Usage

(Optional) c3d features

you can use video-classification-3d-cnn-pytorch to extract features from video.

Steps

1. preprocess videos and labels

```
python prepro_feats.py --output_dir data/feats/resnet152 --model resnet152 --n_frame_steps 40
--gpu 4,5



python prepro_vocab.py
```

1. Training a model

```
python train.py --gpu 0 --epochs 3001 --batch_size 300 --checkpoint_path data/save --feats_dir
data/feats/resnet152 --model S2VTAttModel  --with_c3d 1 --c3d_feats_dir data/feats/c3d_feats
--dim_vid 4096
```

1.

test

2.

opt_info.json will be in same directory as saved model.

3.

```
python eval.py --recover_opt data/save/opt_info.json --saved_model data/save/model_1000.pth
--batch_size 100 --gpu 1
```

TODO

- lstm
- beam search
- reinforcement learning
- dataparallel (broken in pytorch 0.4)

Acknowledgements

Some code refers to ImageCaptioning.pytorch

Pytorch_C3D_Feature_Extractor

pre-trained model (on sport1M) is available:

C3D_sport.pkl

input: video

```
python feature_extractor_vid.py -l 6 -i /data/miayuan/videos/ -o /data/miayuan/c3d_features -gpu
-id 0 -p /data/miayuan/video_list.txt --OUTPUT_NAME c3d_fc6_features.hdf5
```

input: frames

```
python feature_extractor_frm.py -l 6 -i /data/miayuan/frames/ -o /data/miayuan/c3d_features -gpu
-id 0 -p /data/miayua
```