# **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](https://drive.google.com/file/d/1Qi6Gn_l93SzrvmKQQu-drI90L-x8B0ly/view?usp=sharing)
* test-video: [download link](https://drive.google.com/file/d/10fPbEhD-ENVQihrRvKFvxcMzkDlhvf4Q/view?usp=sharing)
* json info of train-video: [download link](https://drive.google.com/file/d/1LcTtsAvfnHhUfHMiI4YkDgN7lF1-_-m7/view?usp=sharing)
* json info of test-video: [download link](https://drive.google.com/file/d/1Kgra0uMKDQssclNZXRLfbj9UQgBv-1YE/view?usp=sharing)

## **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](https://github.com/yunjey/pytorch-tutorial/tree/master/tutorials/03-advanced/image_captioning)

## **Usage**

### **(Optional) c3d features**

you can use [video-classification-3d-cnn-pytorch](https://github.com/kenshohara/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

test

opt\_info.json will be in same directory as saved model.

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](https://github.com/ruotianluo/ImageCaptioning.pytorch)

# **Pytorch\_C3D\_Feature\_Extractor**

pre-trained model (on sport1M) is available:

[C3D\_sport.pkl](http://imagelab.ing.unimore.it/files/c3d_pytorch/c3d.pickle)

### **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