Codes and model of AURL

Descriptions

The source codes are for evaluating our AURL model, including the inference script “main.py”, the network “network.py”, the trained **AURL Model** with $\tau = 0.05$ (i.e., “AURL662_checkpoint.pth.tar”), and the results of top-1 accuracy (i.e., “UCF_top1acc.log” and “HMDB_top1acc.log”) using 1 video clip under 1 test split protocol on the UCF101 and HMDB51 datasets. These top-1 accuracy results are shown in Table 2 (see AURL (ours)) of the manuscript.

You can start the codes from the execute command “run.sh”.

Requirements

```
Python 3.6
Pytorch 1.7.1
torchvision 0.8.2
GoogleNews-vectors-negative300.bin
nltk_data

AURL662_checkpoint.pth.tar
```

Pytorch and torchvision:
pip3 install torch==1.7.1+cu110 torchvision==0.8.2+cu110 -f https://download.pytorch.org/whl/torch_stable.html

GoogleNews-vectors-negative300.bin:
wget https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz -O /workplace/word2vec/GoogleNews-vectors-negative300.bin.gz
gunzip -c /workplace/word2vec/GoogleNews-vectors-negative300.bin.gz > /workplace/word2vec/GoogleNews-vectors-negative300.bin

nltk_data:
pip3 install nltk
python3 -c “import nltk; nltk.download(‘wordnet’)”

Download our **AURL Model**: AURL662_checkpoint.pth.tar

Dataset

Please download related datasets: UCF101, HMDB51

Data folder structure:

```
dataset/
├── class
│    ├──xxx.avi
│    └──other videos
└── other classes
```

Input and Configs

HMDBPath: Input the path of the HMDB51 dataset.
UCFPath: Input the path of the UCF101 dataset.
datasetName: Input "UCF" or "HMDB". The evaluation will be implemented on the corresponding dataset.
clip_len: Number of frames of each sample clip. In the manuscript, we set it as 16.
n_clips: Number of clips per video. In the manuscript, we set it as 1 or 25.
size: Size of the input image. In the manuscript, we set it as 112.
weights: Input the path of the trained AURL model "AURL662_checkpoint.pth.tar".
wordsmodel: Input the path of "GoogleNews-vectors-negative300.bin"
nltkPath: Input the path of "nltk_data".

Output

When predicting the class name of an input video, the inference code will print the result as follow:

```
HMDB Top-1 acc: 27.3721548921076 pred: kiss label: wave
```

"Top-1 acc" is the top-1 accuracy; "pred" is the predicted class name; "label" is the ground-truth. When the implementation is done, the top1 accuracy shown in the final output line is the top1 accuracy on the current dataset.

Test

Please modify the config and run:

```
sh run.sh
```