CSCE689 Project

Name: Pei Chen UIN: 130005135

Main Improvements

1. Remove CRF layer and add several Batch Normalization and Dropout layers in the original CNN layers;
2. Differentiate each frame with the 1st frame within a video because the changes over the whole video is more important to detect the actions and the real content of the video is actually noise;
3. Achieved good performance in all the 3 types of test videos and the result figures look reasonable compared to gold standard.

Research Topic

Detecting foot sentiment in videos. Here are 3 types of foot sentiment.

Nervous Pacing: Many people will pace when they are stressed. This acts as a pacifier, as all repetitive behaviors do.

Foot withdrawing: During situations like interviews, interviewees will suddenly withdraw their feet and tuck them in under their chains when they are asked sensitive questions they might not like.

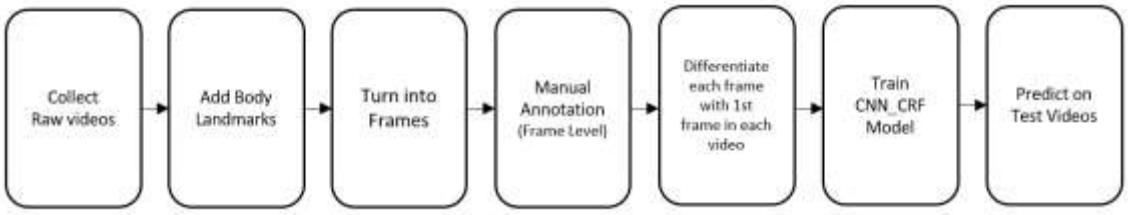
Foot turning away: When we are talking to someone, we might signal that we need to leave by gradually or suddenly pointing one foot toward the door. This is our non-verbal way of communicating "I have to go."

Dataset

I collect 22 videos clips for all the 3 types of actions under different situations for both training and testing. Each clip is about 3 seconds and marked “Body Landmarks” using *OpenPose*. The splitting is as following:

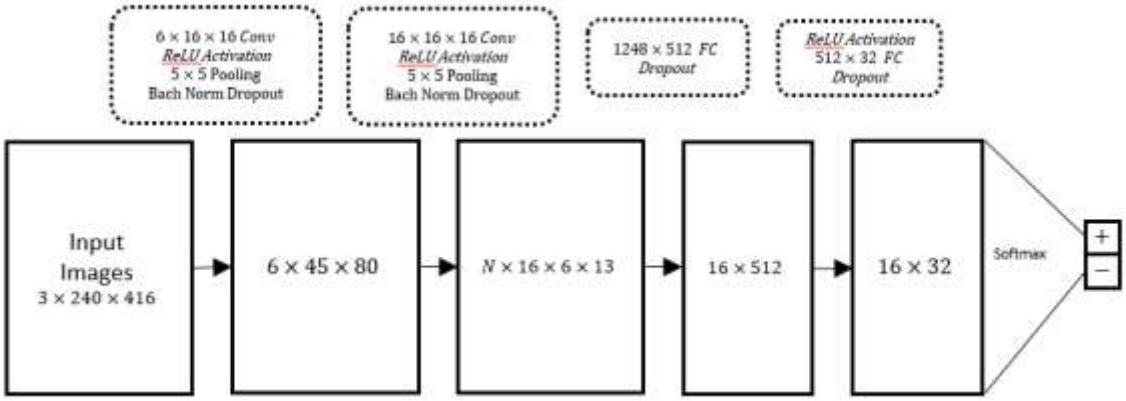
	Training Data	Testing Data
Nervous Pacing	foot_pacing_trian1.mark.mp4 foot_pacing_trian2.mark.mp4 foot_pacing_trian3.mark.mp4 foot_pacing_trian4.mark.mp4 foot_pacing_trian5.mark.mp4 foot_pacing_trian6.mark.mp4	foot_pacing_test1.mark.mp4 foot_pacing_test1.mark.mp4
Foot withdrawing	foot_withdrawing_trian1.mark.mp4 foot_withdrawing_trian2.mark.mp4 foot_withdrawing_trian3.mark.mp4 foot_withdrawing_trian4.mark.mp4 foot_withdrawing_trian5.mark.mp4 foot_withdrawing_trian6.mark.mp4	foot_withdrawing_test1.mark.mp4 foot_withdrawing_test2.mark.mp4 foot_withdrawing_test3.mark.mp4
Foot turning away	foot_turning_away_train1.mark foot_turning_away_train2.mark foot_turning_away_train3.mark foot_turning_away_train4.mark	foot_turning_away_test.mark.mp4

Procedures



1. Collet videos from YouTube and trim them into 3 seconds per video.
2. Use *OpenPose* tools mark “Body Landmarks” in all clips;
3. Then transform the videos into frames, and store each frame as an image;
4. After that, manually annotate each frame as Positive (with target action) or Negative (without target action);
5. Differentiate each frame with 1st frame in each video;
6. Train separate CNN models for each action type to classify the frames;
7. Test on the corresponding test data and get the scores for each frame.

Architecture:



```
class Net(nn.Module):
    def __init__(self):
        super(Net, self).__init__()
        self.conv1 = nn.Conv2d(3, 6, 16)
        self.pool = nn.MaxPool2d(5, 5)
        self.conv2 = nn.Conv2d(6, 16, 16)
        self.fc1 = nn.Linear(16 * 6 * 13, 512)
        self.fc2 = nn.Linear(512, 32)
        self.fc3 = nn.Linear(32, 2)
        self.drop_out = nn.Dropout(0.5)
        self.BN1 = nn.BatchNorm2d(6)
        self.BN2 = nn.BatchNorm2d(16)

    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x)))
        x = self.drop_out(self.BN1(x))
        x = self.pool(F.relu(self.conv2(x)))
        x = self.drop_out(self.BN2(x))
        x = x.view(-1, 16 * 6 * 13)
        x = F.relu(self.fc1(x))
        x = self.drop_out(x)
        x = F.relu(self.fc2(x))
        x = self.drop_out(x)
        x = self.fc3(x)
        return x
```

The main architecture is a 2-layer CNN (Convolutional Neural Network) followed by 3-layer Feedforward Neural Network (FNN). I make some changes compared to previous model: I added dropout layer for all CNNs and FNNs and Bach Normalization for the CNN.

Another trick I use is to differentiate each frame with the 1st frame within a video. Intuitively, the changes over the whole video is more important to detect the actions and the real content of the video is actually noise.

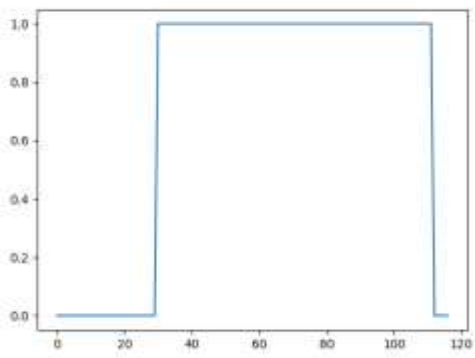
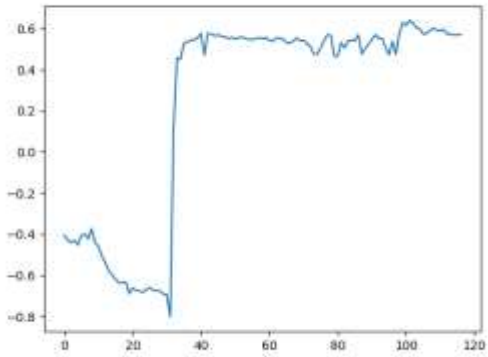
I tried differentiating each frame with the previous frame within a video, but the performance is not as good.

Input

In order to maintain the consistency for all the clips, I turn all the frames into shape of (240, 416, 3).

Output

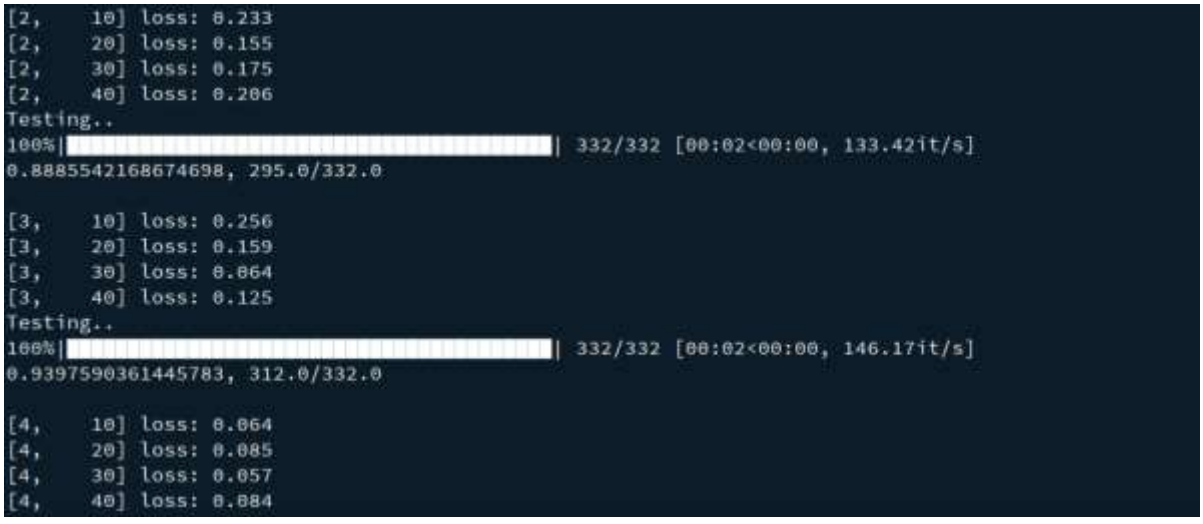
For each test clip, I predict the score of each frame with the trained model and store them in a corresponding *JSON* file. The following figure are the gold labels and predicted results from foot_pacing_test1 data.

Hyperparameters

Batch_size	Optimization	Learning Rate	Max_epoch
16	SGD	0.001	30

Training and Testing Performance



Performance	Foot pacing	Foot withdrawing	Foot turning away
Accuracy	0.9412	0.8230	0.6835

Next Steps:

1. Improve the model to address the challenging foot withdrawing and foot turning away detections.
2. Use tricks such as data reinforcement and transfer learning to further performance.

Links

Videos: https://www.youtube.com/playlist?list=PLlUqXrHW9mC_GlGbgiAUmPIwDMZWuLSP4

Code: https://github.com/brickee/foot_sentiment.git