Sapienza University of Rome

Convolutional 3D
Body Movement Recognition

Ibis Prevedello, Jean-Pierre Richa

October 29, 2018

1 Introduction

The goal of this project is to train a model based on Convolutional 3D and Openpose in Tensorflow to recognize body movements from videos. In order to execute this task, first the person's pose is recognized using Openpose and then it is used to recognize the body member movement.

The dataset used for the training is a subset of the Human3.6M Dataset, these videos were first segmented by the course students and divided in a total of 52 classes.

In order to minimize the training time, the dataset was first converted to tfrecords and then used for training the network. All this process was executed using Google Compute Engine instance running a Tesla K80 GPU.

2 Dataset Augmentation

The dataset segmentation, composed of 52 classes, was given in two json files, one for the training and one for the test. Analyzing the number of samples for each category was noticed a big difference from class to class.

In order to have a balanced dataset, was decided to extract a subset of 26 classes with the highest number of samples, resulting in the following list:

- Head
  - Turn right
  - Turn left
  - Raise

  - Lean forward

- Right/Left Arm
  - Shoulder extension
  - Shoulder adduction
  - Shoulder flexion
  - Shoulder abduction
  - Elbow flexion
  - Elbow extension
  - Roll the wrist

- Right/Left Leg
  - Hip flexion
  - Hip extension
  - Knee flexion
  - Knee extension

For the training dataset, after having a subset of the categories, it needed to be augmented, in order to achieve that the idea was to flip the images from the arm and leg classes and add them to the opposite member class. For example, if we flip the right shoulder extension we obtain a left shoulder extension.

Having that, we obtained a total of 400 samples for each class, with a total of 10.400 samples for the training set and 1.900 samples for the testing set.

3 Network Training

The network was trained in two different ways. The first was to transfer the learning from an already trained network used to recognize sports activities to our task, an the second one to train the network from scratch. Both networks were trained for 10 epochs and using a batch of 10 samples.

Using transfer learning, the last layer was deleted and added a new one with the number of classes of our task. During the first 3 epochs only the last layer weights were updated with a learning rate of 1×10^{-3} and after that the whole network was fine tuned with a learning rate of 1×10^{-4}.

For the second approach no pre-trained model was used and the whole network was trained using a learning rate of 1×10^{-3}.

The training and the testing set was evaluated after each epoch and it can be seen using Tensorboard.

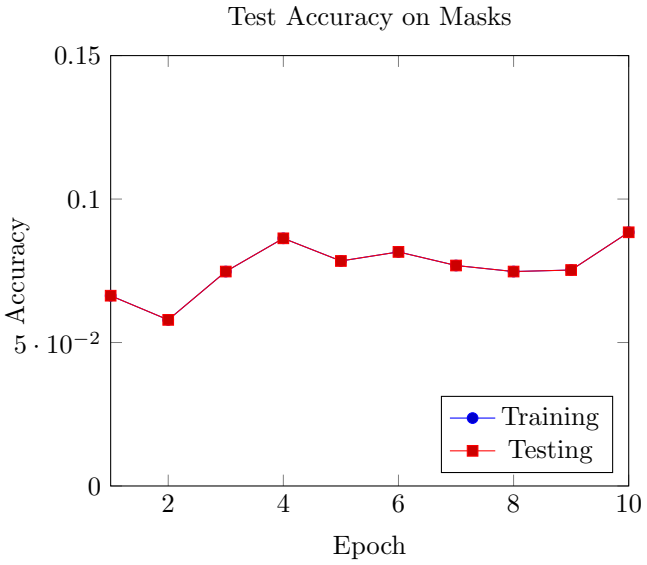

Figure 1: The plot shows the how the test accuracy varies for the dataset Masks changing the number of hidden nodes in the LSTM.

4 Results

It was decided to train both networks for only 10 epochs because it is enough to see how the accuracy is improving over time, allowing to compare the two strategies (transfer learning and training from scratch) and also because the dataset is quite extensive and its training takes a long time, and because it is being trained on the cloud and it is payed by hour.

The data presented here was saved during the trained and collected from Tensorboard.

5 Implementation

The project has been implemented in Python using Tensorflow. Below, the list of Python files implemented for the project with a brief description of their behavior.

- **generate_tfrecords.py:** generating the tfrecord files used for the training and evaluation of the network.
- **train.py:** train the network specifying parameters and the dataset.
- **activities.py:** list of activities to be trained and the number of samples per activity to augment data.

- **c3d_model.py:** C3D model implementation.

Extra file:

- **pose_list.py:** shows the list of activities and the frequency of activities chosen to the training.

6 Conclusion