

Chapter 1

Classification stage

1.1 Description

After getting all proposed tubes, it's time to do classification. As classifiers we use 3 approaches

- A Recursive Neural Network (RNN) Classifier
- A Linear Classifier
- A Support Vector Machine (SVM) Classifier

The general structure of the whole network is depicted in figure 1.1

1.2 RNN Classifier

First approach includes a RNN as classifier. The reason is that we do not have

1.3 Linear Classifier

Using a Linear classification we follow the next steps:

1. Firstly we extract the candidate tubes for the whole video and get their corresponding features.
2. For each candidate tube, we get a tensor with dimensions $(\#clips, 256, \#frames, 7, 7)$ where $\#frames$ is the number of frames of each clip. So we calculate the average value for all the clips and we get a tensor with dimension $(256, \#frames, 7, 7)$.
3. Our linear Classifier gets as input the previous tensor.

In order to consider a valide classification, we set as confidence threshold equal to 0.5, 0.75 and 0.9 . We firstly examine our classifier's performance only for jHMDB dataset. In figure ?? we can see the first results which are disappointed.

1.3.1 Improving results

1.4 Final Improvements

After classification, we relize that a lot of classified tubes overlap and represent the same action. So, we use again NMS algorithm in order to remove unnecessary tubes. The new model can be seen in figure 1.2.

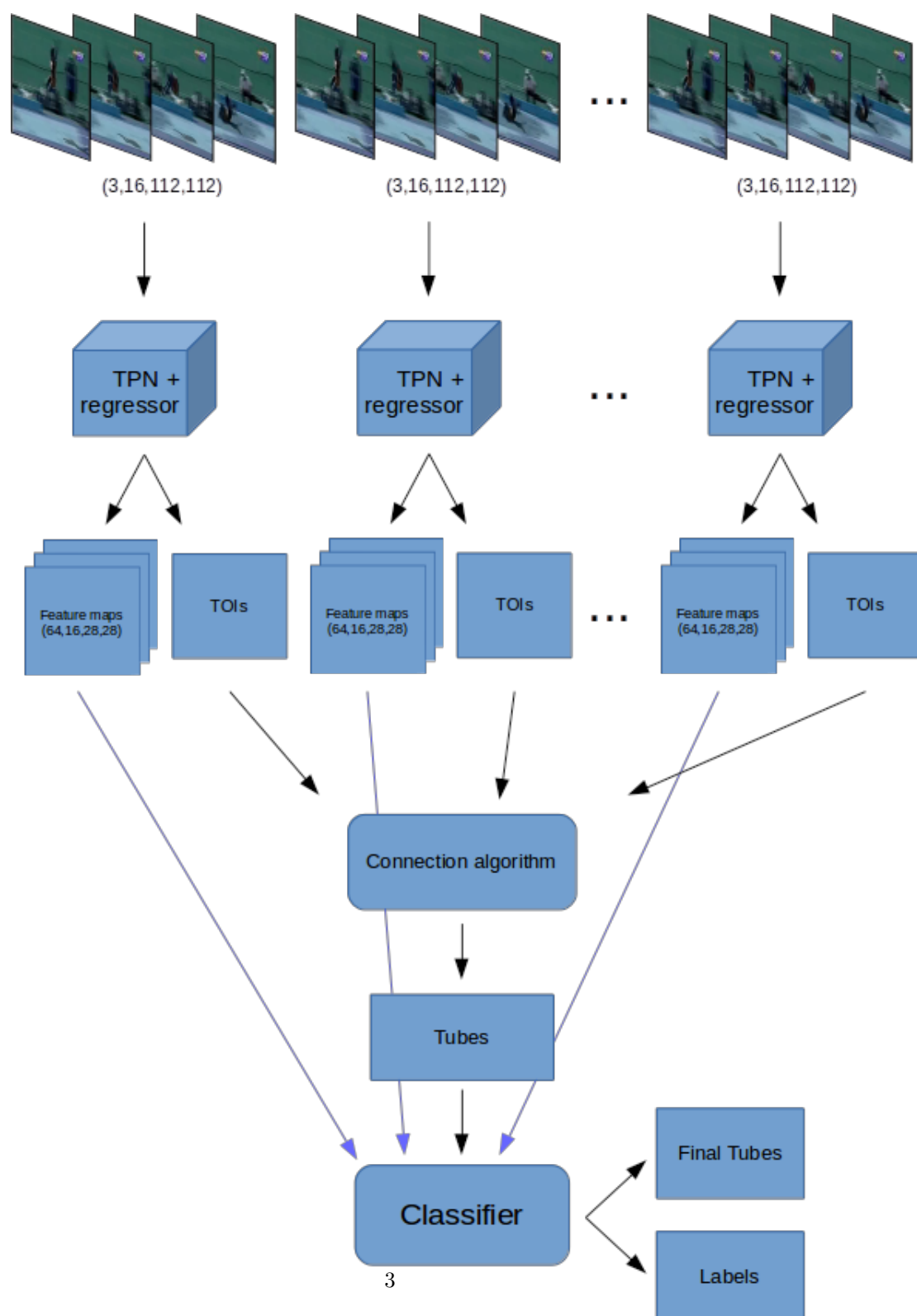


Figure 1.1: Structure of the whole network

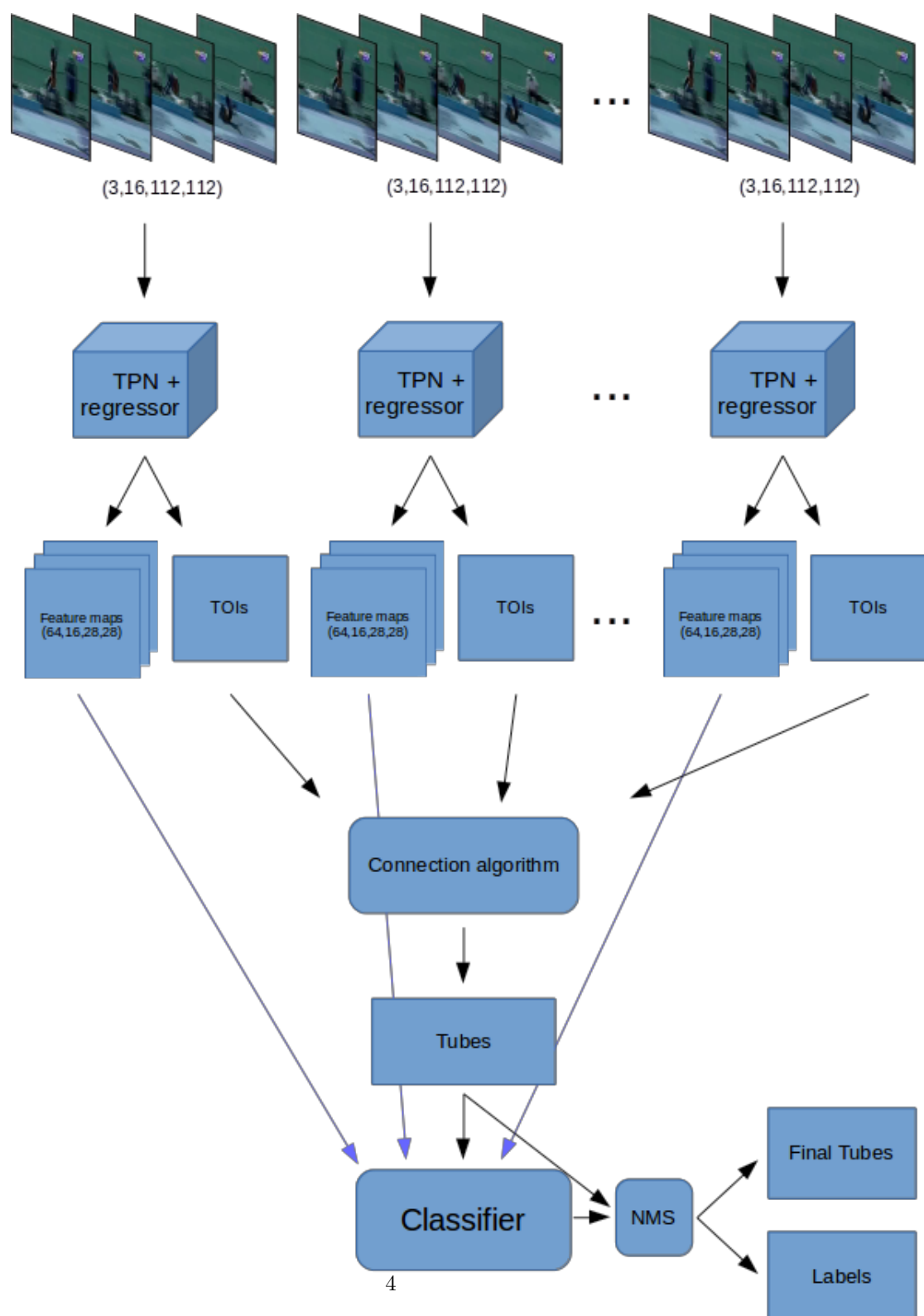


Figure 1.2: Structure of the network with NMS