華南理工大學

South China University of Technology

The Experiment Report of Machine Learning

School: School of Software Engineering

Subject: Software Engineering

Author:
Wanghua Shi
Wenjun Liang
Weiwen Hu

Supervisor:
Mingkui Tan or Qingyao Wu

Student ID:
201630676843
201630664963
201630676713

Grade:
Undergraduate

December 28, 2018

Face Detection Based on Neural Network

Abstract—Based on reading and understanding the principles of the given paper «Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks», we run the given codes with well trained network model on sixty-four test pictures. Finally, we got 64 pictures which are well marked with a rectangular box and five facial landmarks positions.

I. Introduction

Face detection and alignment are essential to many face applications, such as face recognition and facial expression analysis. However, the large visual variations of faces, such as occlusions, large pose variations and extreme lightings, impose great challenges for these tasks in real world applications. Besides, most of previous face detection and face alignment methods ignore the inherent correlation between these two tasks. Though several existing works attempt to jointly solve them, there are still limitations in these works. On the other hand, mining hard samples in training is critical to strengthen the power of detector. Nevertheless, traditional hard sample mining usually performs in an offline manner, which significantly increases the manual operations. To solving these tasks discussed above, we learned and adopted the unified cascaded CNN proposed in the paper [1].

II. Methods and Theory

A. Overall Framework

The proposed Cascaded Convolutional Networks consist of three stages. A brief description is given below.

1) Produce candidate windows quickly through a shallow CNN: Exploit a fully convolutional network, called Proposal Network (P-Net), to obtain the candidate facial windows and their bounding box regression vectors. Employ non-maximum suppression (NMS) to merge highly overlapped candidates.
2) Refines the windows by rejecting a large number of non-faces windows through a more complex CNN called Refine Network (R-Net). Performs calibration with bounding box regression and conducts NMS again.
3) Uses a more powerful CNN to refine the result again and output five facial landmarks' positions.

B. CNN Architectures Improvements

By analyzing the limitations of multiple CNNs in [2], reduce the number of filters and change the 5×5 filter to 3×3 filter to reduce the computing while increasing the depth to get better performance. After the convolution and fully connection layers (except output layers), apply PReLU [3] as non-linearity activation function.

C. Training Model Selection

- Train CNN Detectors
  Three tasks of training CNN detectors are leveraged: face/non-face classification, bounding box regression, and facial landmark localization. Details are summarized below.
  1) Face Classification: Use the cross-entropy loss:
  $$L_i^{\det} = -(y_i^{\det}\log(p_i) + (1 - y_i^{\det})(1 - \log(p_i))) \tag{1}$$
  where p_i is the probability produced by the network that indicates sample x_i being a face. The notation $y_i^{\det} \in \{0, 1\}$ denotes the ground-truth label.
  2) Bounding box regression: For each candidate window, predict the offset between it and the nearest ground truth, employ the Euclidean loss for each sample x_i.
  3) Facial landmark localization: Still use the Euclidean loss and minimize it. Since there are five facial landmarks, $y_i^{landmark} \in \mathbf{R}^{10}$.
- Multi-Source Training
  Use a sample type indicator to handle different types of training images in the learning process. The overall learning target can be formulated as:
  $$\min \sum_{i=1}^{N} \sum\nolimits_{j\in\{\det, box, landmark\}} \alpha_j \beta_i^j L_i^j \tag{2}$$
  where the N is the number of training samples and α_j denotes on the task importance. $\beta_i^j \in \{0, 1\}$ is the sample type indicator.
- Online Hard Sample Mining
  [1] conducts online hard sample mining which is adaptive to the training process. In each mini-batch, there are two steps:
  1) Sort the losses computed in the forward propagation from all samples and select the 70% of them as hard samples.
  2) Only compute the gradients from these hard samples in the backward propagation.

III. Experiments

A. Dataset

1) Use WiderFace for face classification and face bounding box regression when training PNet, RNet and ONet.
2) Use Training Data Set for face feature point regression.

B. Environment for Experiment

- Anaconda3
- pytorch 0.4.1
- opencv-python
- tensorflow(only for python 3.4,3.5,3.6)

C. Implementation

- Read the paper about MTCNN [1] We have printed the five-page paper, read it carefully and summarized it in the part II.
- Run the given codes
  1) Get the complete codes in MTCNN_pytorch.
  2) Install the environment (Details Omitted).
  3) Test the given models: Run the file "test_image.py"(Use the command below), and examined the results in the path ".../mtcnn_pytorch/data/you_result/".

```
cd mtcnn_pytorch/
python test_image.py
```

We have found all the faces of test images have been framed in a rectangle and five facial landmarks positions of faces (left eye, right eye, nose, left mouth corner and right mouth corner) have been marked with red points. Besides, the generating speed of trained model is fast.

IV. Conclusion

In this experiment, we have adopted a multi-task CNNs based framework for joint face detection and alignment. We made our efforts to understand the basic theory of face detection using neural network, understand the processes of MTCNN and use it in practice. Eventually, we have a basic understanding of CNNs, online hard samples mining strategy and joint face alignment learning.

References

[1] Kaipeng Zhang. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, 2016.

[2] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection" in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5325-5334.

[3] K. He, X. Zhang, S. Ren, J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification" in IEEE International Conference on Computer Vision, 2015, pp. 1026-1034.

Appendix A
Task Division Table in This Experiment

Group Member	Main Task
Wanghua Shi	Read paper, run codes and write experiment report
Wenjun Liang	Read paper and understand model
Weiwen Hu	Read codes and understand the algorithm