Face Size Detection using Deep Learning

Batkhongor Tseden*, Hae-Yeoun Lee*
*Dept of Software Engineering, Kumoh National Institute of Technology

Abstract

Many deep learning approaches are studied for face detection in these days. However, there is still a performance problem to run efficiently on devices with limited resources. Our method can enhance the detection speed by decreasing the number of scaling for detection methods that use many different scaling per image to detect the different size of faces. Also, we keep our deep learning model easy to implement and small as possible. Moreover, it can be used for other special object detection problems but not only for face detection.

1. Introduction

Detecting faces in different sizes is one of the most important part in every single face detection methods. Even there is only one big sized face, current some methods check a number of different sizes of faces [1]. It spends resource for unnecessary operations and increase detection time. As shown in Figure 1, DDFD [2], for example, scales image 10 times from bigger to smaller.

![Figure 1. Detected faces using DDFD [2]](image)

Our method is trying to solve that problem by finding face size and it is possible to use minimal number of scaling for face detection. Moreover, Figure 2 shows heat-map on each face detection scales, here we can see that there is higher probability to detect wrong face at some scales. Therefore, if we find right sizes, it is able to decrease false positive face detection. First of all, we proposed to find minimum face size in this paper.

![Figure 2. heap-map at scale 6.2 and 3.15, respectively on Figure 1](image)

2. Proposed algorithm to detect face size

We created training data using AFLW dataset [3] images for image scaling and labeled images between 0 - 4 according to proportion of minimum face size and image size as shown in Figure 3.

![Figure 3. Label calculating for training images](image)
We used Alexnet [4] as deep learning model and Caffe [5] as library for our face size detection method. Then, we trained Alexnet and modified Alexnet which has same structure but less parameters.

3. Experimental results and conclusion

Using PASCAL Face [6], AFW [7] datasets, both models were tested on NVIDIA GeForce GTX 1080 Ti. The accuracy analysis result is shown in Table 1.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Model size</th>
<th>Images per second</th>
<th>Accuracy on Pascal [7]</th>
<th>Accuracy on AFW [8]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexnet 222Mb</td>
<td>5</td>
<td>86%</td>
<td></td>
<td>48%</td>
</tr>
<tr>
<td>Modified Alexnet</td>
<td>40Mb</td>
<td>12</td>
<td>85%</td>
<td>48%</td>
</tr>
</tbody>
</table>

Table 1. Comparison of 2 trained models

To use this method as part of face detection, its accuracy must be more than 90 percent, but it could not perform like we expected. However, there were only 18,000 images for training which is not enough to perform excellent. We need to increase number of images on training by combining AFLW [3] and FDDB [8] datasets to improve accuracy. Furthermore, there is need to use other methods like subsampling to increase the number of images.

Acknowledgement

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1 A1B03030432).

References