Face Detection and Facial Attribute Editing

Mohammadamin Ahantab
amin_ahantab@mytum.de

Mohammadamin Barekatain
m.barekatain@tum.de

Azade Farshad
azade.farshad@tum.de

Yasaman Rajaee
yasaman.rajaee@tum.de

1. Introduction

Face detection is the crucial step of many subsequent face-related applications, such as face alignment, recognition, verification, tracking, key point detection, pose estimation, etc. It has been well developed over the past few decades and has been an active research area with many successful traditional and deep learning methods.

Facial attribute editing, aims to manipulate a face image by controlling the facial attributes of interest (e.g. gender, expression, mustache, and age). Learning controllable attributes from images is fundamental to a number of real-world applications in computer vision, such as image generation and image transformation.

1.1. Problem Statement

Since it is practically infeasible to collect images with arbitrarily specified attributes for each person, we are interested in the problem of manipulating natural images by controlling some attributes of interest. The current approaches take as an input the cropped face and then perform facial attribute editing task. However, in the real-world scenario, the input is not a cropped face. Hence, face detection is essential for facial attribute editing in real-world scenario.

1.2. Related Works

Over the past years, there has been extensive works in the face detection such as SSH[8], S3FD[10] and FAN[9]. Moreover, there exists many approaches for image attribute controlling area such as Fader Networks[6] and BiGAN[5]. Our proposed method has 2 contribution: 1) adopting existing architecture of object detection for face detection task 2) synthesizing face detection and existing facial attribute editing models.

2. Dataset

- WIDER FACE: includes 32k images and 393k labeled faces[4].
- FDDB: contains the annotations for 5171 faces in a set of 2845 images taken from the Faces in the Wild data set[2].
- AFLW: a large-scale collection of 25k annotated face images gathered from the web, providing a large variety in appearance as well as general imaging and environmental conditions.
- CelebA: consists of more than 200K celebrity images, each with 40 attribute annotations which belong to more than 10k identities[1].
- Datasets such as UMD faces may also be used[3].

These datasets include colored two dimensional images of human faces. They provide bounding boxes for human face which are required for face detection. The input data in this project is images and the output would be bounding boxes around every face in the input image.

3. Methodology

This project consists of three phases: In the first phase we tackle the problem of face detection by adopting and fine-tuning Single Shot MultiBox Detector (SSD)[7] which has been very successful for object detection task.

In the second phase we first train Fader Networks[6] on CelebA dataset. After training, the model can generate different realistic versions of an input image by varying the attribute values.

In the last phase, the pre-trained detection model and Fader Networks are used jointly to perform facial attribute editing of the detected faces. This is done by first running the face detection to extract faces from the input image, then resizing these extracted regions and then feeding them to the Fadar network.

Our group has access to GPU computing power on servers and a laptop with Nvidia Quadro P4000(8GB) GPU.

4. Outcome

A pipe-line which can detect the faces and then change its facial attributes.

References

[1] CelebA dataset. http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html. Accessed: 2017-12-20. 1

[2] FDDB dataset. http://vis-www.cs.umass.edu/fddb/. Accessed: 2017-12-20. 1

[3] UMD dataset. (http://www.umdfaces.io/). Accessed: 2017-12-20. 1

[4] WIDER FACE dataset. http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace. Accessed: 2017-12-20. 1

[5] J. Donahue, P. Krähenbühl, and T. Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016. 1

[6] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato. Fader networks: Manipulating images by sliding attributes. *arXiv preprint arXiv:1706.00409*, 2017. 1

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 1

[8] M. Najibi, P. Samangouei, R. Chellappa, and L. Davis. Ssh: Single stage headless face detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4875–4884, 2017. 1

[9] J. Zhang, X. Wu, J. Zhu, and S. C. Hoi. Feature agglomeration networks for single stage face detection. *arXiv preprint arXiv:1712.00721*, 2017. 1

[10] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li. S3fd: Single shot scale-invariant face detector. *arXiv preprint arXiv:1708.05237*, 2017. 1