Project No. VI20172020068

Tools and Methods for Video and Image Processing to Improve Effectivity of Rescue and Security Services Operations (VRASSEO)

Detection of facial key points

Technical report

Kanich O., Goldmann T., Drahanský M.

Brno University of Technology
Faculty of Information Technology
Božetěchova 1
Brno 612 66, Czech Republic

December 2019

Contents

Abstract

This report describes an algorithm for detection of 98 facial key points using convolution neural nets (CNN). The neural nets use novel design based on state of the art design for similar tasks. WFLW (Wider Facial Landmarks in-the-wild) database which contains 10,000 face images in atypical poses (illumination, makeup, expression, occlusion, etc.) has been used for training and testing.

1 Introduction

Facial recognition is trending method for identification or verification of people. Face features are one of the most used biometric characteristics nowadays. The rise of these methods is closely related to the achievements of deep (convolutional) neural networks (CNN or DCNN). These greatly increased success rate of facial recognition.

Nevertheless, these methods still struggle when face is not positioned forwardly to the camera. Methods which are trying to cope with unusual face orientation are based on the *facial key points* (FKP) [1]. FKP are usually set-up, so they approximate position of important parts of face (such as eyes, nose, lips, etc.). Small rotation or tilt of head is not a problem – FKP are in similar positions and their mutual relation are similar as well. As one can image, using side portrait as an input image for facial recognition makes huge difference. Possibilities of different face orientation also opens the topic of 3D face recognition.

This work is trying to make a first step of solving some of these issues. The goal is to define and detect great number of FKP. When designing solution, one has to have in mind that FKP should be later processed so that rotation and tilt of the head could be estimated or to be used for facial recognition alone. Another obstacle in realistic applications are face expression, illumination, makeup and other "defects" present in face images. To test a solution against these "flawed" images appropriate database (called usually in-the-wild) should be used. In this work part of *Wider Facial Landmarks in-the-wild* (WFLF) database is used for validation.

First section of the work is dealing with definition and state-of-the-art in scope of FKP. Second part is describing chosen database for testing of the designed solution. Third section is focusing on the core of this work – the FKP detector. Last part is discussing achieved results and further improvement of the presented method.

2 Facial key points

Preliminary information about facial key points or facial landmarks could be found in introduction section. There are many possible definitions of FKP in face. For example, the method chosen for forensic 3D facial identification (FIDENTIS [2]) is used. In this tool, FKPs are defined in the following locations: eyes (6), nose (2), lips (5), chin (2) and forehead (1). Example can be seen on Figure 1. FKP position can describe the face but also define position of face, nose, eyes, and some changes based on mimic muscles movement. Because of that these points can be used for pose recognition, emotion recognition, face dysmorphia and other tasks [3]. Some models also define FKP on eyebrows. A precise definition depends on intended use – eyebrows could be more important for emotion, great number of bilateral points for dysmorphism and so on. It is certain that these 16 points are pretty discriminative for straight portrait image type. They form double cross on the face which express the face nicely.

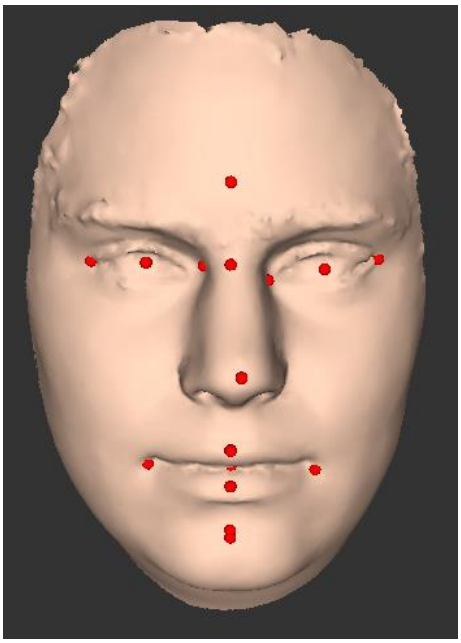

Figure 1: Sample image of basic facial key points (taken from [2]).

However the intended use is the *face pose estimation* and working with rotated or tilted face images. For this purpose the border points are also needed. That means an increased number of points around chin, generally in the edges of face and ideally adding some points around the ear. These will probably not be seen in frontal view, but they will help in case of side portrait images. While the definition of points that perfectly suit the work objective is important, there are other essential questions: Which database (dataset) will be used for testing? Who will annotate the database images based on tailor-made FKP?

It is planned to use DCNN for FKP detection – in this case a lot of data is needed, and these questions are more important than a perfect definition of points. Dataset is described in the next section, but the points will be described here and they were finally chosen based on the dataset. 98 FKP are used so that: 33 points belong to edge of the face (edge of chin and cheeks), eyebrows are described with 18 points, 9 points define nose, 18 points belong to the eyes and finally 20 points describe lips. Exact position and numbering of points can be seen in Figure 2.

Figure 3 shows two example images of head rotation to the side portrait position. As can be seen in this figure, there are still a lot of FKP visible. It should be noticed that curve created by blue points (describing chin and cheeks) is changing dramatically its shape, as well as getting close to the black points (nose definition), what can be used for face pose estimation. Enormously increased number of FKP should help if part of the face will not be visible or if some of the points are not detected in the right spots.

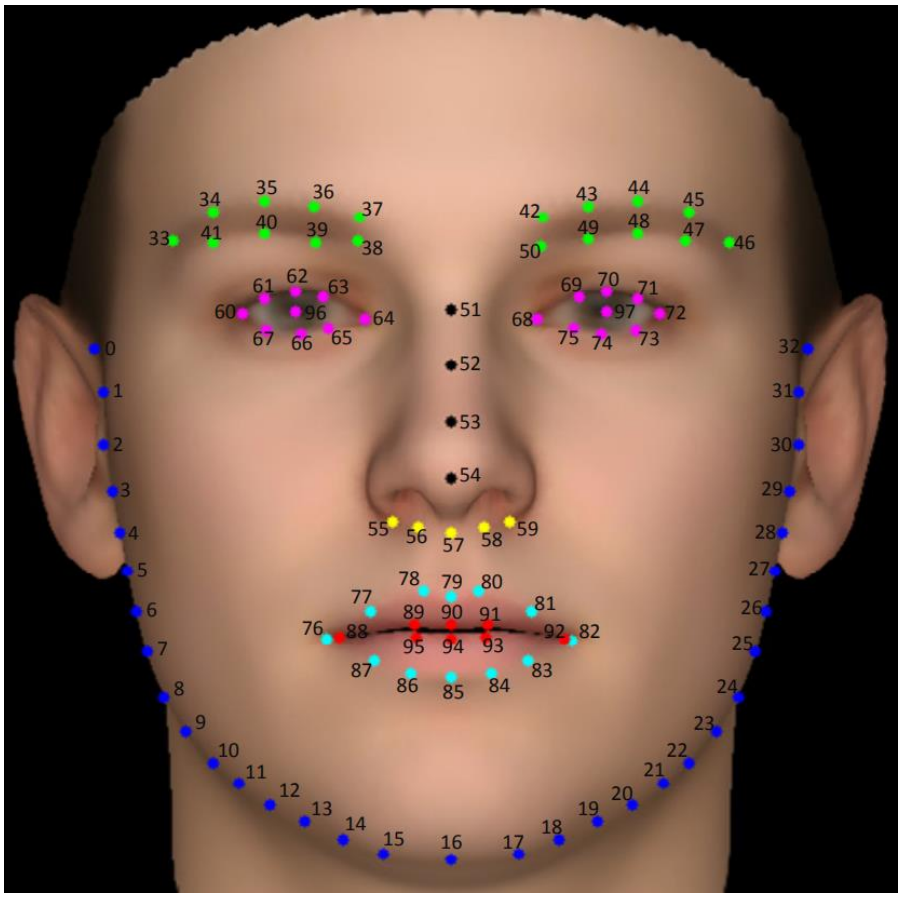

Figure 2: Definition of facial key points use in this work (taken from [4]).

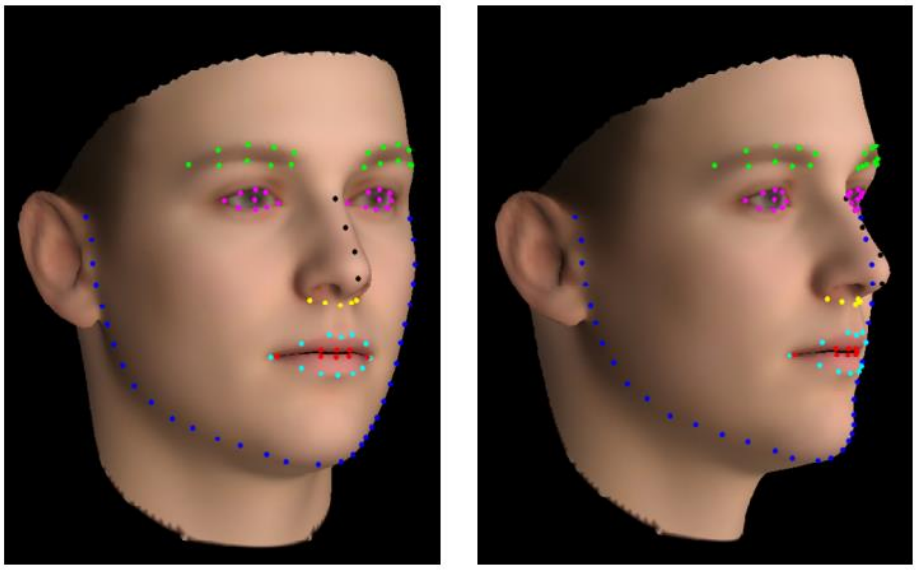

Figure 3: Chosen facial key points in different face rotation (edited from [4]).

3 Database WFLF (Wider Facial Landmarks in-the-wild)

One of the goals of this work is to allow the usage in realistic scenarios. It means that it cannot be assumed that the position of the face will be perfect, or that the quality of the image will meet some common standards. It is not meant the quality in the scope of used camera or resolution but in the scope of user cooperation. Nowadays, for these purposes the special databases called "in-the-wild" are created. These databases are acquired (or edited) from real videos, when persons are not instructed, nor they follow any acquisition instruction.

The fact is that success rate of face recognition drops rapidly if it is not prepared for these kinds of images. Because of that there are several available databases of face-in-the-wild. As it has been mentioned in the previous section, in this work Wider Facial Landmarks in-the-wild database is used [4][5] because it contains problematic images and also all of these images are manually annotated with 98 FKP (as has been described in the previous section). Samples of images can be seen in Figure 4. Each row of this figure also shows one type of "defects" present in the database. These are large pose (first column), expression (second column), unusual illumination (third column), (excessive) usage of makeup (fourth column), occlusion (fifth column) and blur (sixth column).

Figure 4: Sample images from database WFLF (taken from [4]).

Overall, the database contains 10,000 face images with these 98 FKP annotations, attribute annotation (types of "defects") which allows better understanding what is going on with the tested algorithm.

4 Key points detector

CNN are the most often used tools in facial recognition. Detailed description of neural networks and CNNs is in literature (and even in previous technical report) such as [6].

FKP detection is not a new topic in facial detection. At the beginning of the design of a new detector the deep literature recherché has been done. Two important articles were found which employ similar solutions. First one [3] uses high number of FKP but the used methods are just basic. The whole article is more like a tutorial material. Second article [1] uses more profound methods but detects only 15 FKP. The chosen CNN and methods should use what is useful for the solution from both articles.

Firstly, processing of images from WFLF database will be discussed. This step is closely linked with data augmentation. After that CNN model for 98 FKP detection will be described.

4.1 Preprocessing and data augmentation

Neural networks require standardized size of images. That is the first step to determine the suitable size. If the image is too small, then there is high possibility that some FKP would be out of the image. On the other hand, bigger size implies bigger CNN and longer time to train the detector. In [1] the mentioned approach uses the size of 98 × 98 pixels. Analysis of figures in [1] shows that the face barely fits this size and very often there are missing points on chin and cheeks. This resolution is not sufficient, in [3] the image size is defined as 224 × 224 pixels what looks more realistic. Because there are used more FKP in this work, it was decided to slightly enlarge the size to 288 × 288 pixels. Before changing the size of the images in database to a desired size there is one important step necessary – it is the conversion of the image to the greyscale. Other color channels would add more complexity (dimension) to the CNN.

There are several ways how to change the size of images in the database. It can be just a cropped part of the image, scaling of the whole image, etc. Usually, data augmentation methods are used for that. This means that from one image in a database several images for CNN training are generated. The possibilities are: **Whole** image **scaling** – where the whole image is slightly down- or up-scaled and after that crop methods are used. **Flipping** – horizontal and/or vertical flip of the whole image. **Cropping** – where a part of the image of desired size is cropped from the whole image. **Cropped** image **scaling** – where cropped part is scaled and then cropped again. Coordinates of annotated data has to be equally changed (scaled, flipped or cropped). In this work only cropping is used (but other methods are mentioned as the possibility of future extension). Distance of edge from the region of interest (referred in dataset annotation) is set to create cropped image. The larger the padding, the smaller the size of the actual face in the image. That is because width of the image must maintain constant.

4.2 Model of CNN

Proposed model of the CNN is mainly based on the model of NamishNet [1], but with one important change. NamishNet was designed so that it finds only one point, result of the net is one set of coordinates. The idea was that there should be 15 CNNs running to get all 15 FKP. This is hardly imaginable with more FKP, last set of layers had to be changed. The architecture of the proposed net can be seen in Table 1.

Table 1: Architecture of the proposed CNN model for FKP detector.

#	Layer type	Filters	Kernel s.	Other
1	Convolution 2D	16	(5, 5)	Padding (same)
2	Activation			Function (ReLu)
3	Convolution 2D	16	(5, 5)	
4	Activation			Function (ReLu)
5	Convolution 2D	16	(3, 3)	
6	Activation			Function (ReLu)
7	MaxPooling 2D		(5, 5)	Strides (2, 2) Padding (valid)
8	Convolution 2D	32	(3, 3)	
9	Activation			Function (ReLu)
10	Convolution 2D	32	(3, 3)	
11	Activation			Function (ReLu)
12	Convolution 2D	32	(3, 3)	
13	Activation			Function (ReLu)
14	MaxPooling 2D		(3, 3)	Strides (2, 2) Padding (valid)
15	Convolution 2D	64	(3, 3)	
16	Activation			Function (ReLu)
17	Convolution 2D	64	(3, 3)	
18	Activation			Function (ReLu)
19	Convolution 2D	64	(3, 3)	
20	Activation			Function (ReLu)
21	MaxPooling 2D		(3, 3)	Strides (2, 2) Padding (valid)
22	Convolution 2D	128	(3, 3)	
23	Activation			Function (ReLu)
24	Convolution 2D	128	(3, 3)	
25	Activation			Function (ReLu)
26	Convolution 2D	128	(3, 3)	
27	Activation			Function (ReLu)
28	MaxPooling 2D		(3, 3)	Strides (2, 2) Padding (valid)

29	Flatten			
30	Dropout			Rate (0.2)
31	Dense	392	Function (ReLu) Regularizer L2 (0.001)	
32	Dropout			Rate (0.2)
33	Dense	196		

As it can be surely noticed that it is not the only change in the architecture. Over the time, design of CNN model changed and deviated more from the original model. Proposed architecture gets good results on the small test batches and it is considered as final in the time of writing of this technical report.

The model can be divided into five parts: first of them is taking input image and using convolutional layers with small number of filters (16) but rather high kernel size (5 and 3); after that it is down-sampled by maxPooling layer (also quite significantly with pool size 5); second to fourth layer are similar, settings are more usual (kernel size 3) and the only thing which changes is the amount of filters used in convolutional layers (from 32 to 128); all these layers in four parts are alternated with activation layers which use ReLu function; fifth layer is flattening from 3D to 1D after that there are dropout and dense layers, which finally ends in 196 values (98 FKP times two coordinates).

5 Results

The most important thing about the results of proposed solution is that the work is still in progress. Architecture, settings and overall approach to the solution was set, preliminary results look promising, but that do not necessarily mean that the solution will not change in the following optimization period. Unfortunately, there are no final statistical results available. What can be shown are images, which have been used for testing. Figure 5 shows sample image (which was not in the

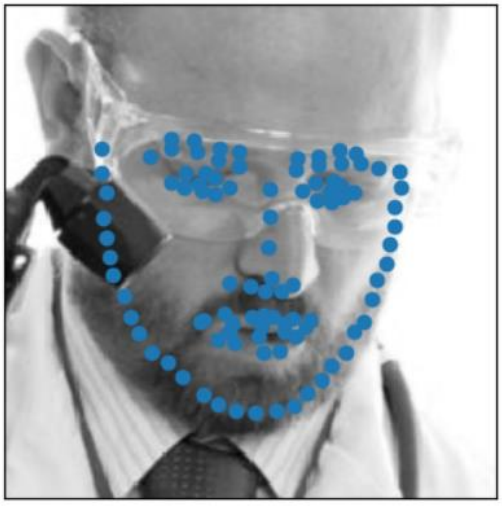

Figure 5: Sample image with detected key points.

training set of the CNN). Image itself shows a lot of "defects" present in the image. There is beard, glasses on the face and some object which occludes the face. Head is also a little bit tilted downwards. Even though the nose points are slightly off the central part of the nose, the overall structure of other points is pretty close their optimal location.

Figure 6 shows more samples which are focused more on the rotation of the head, what is one the main focuses of the presented detector. If the images are numbered from left to right, top to bottom, there are very promising results on the third image, where head of the person is basically in the side portrait position and the points are clearly following the edges of the face. Similarly, astounding results are shown in fourth image where head is rotated in other axis and points are on spots. On the other hand, the last, ninth image is showing person which is looking upwards and is very close to the side portrait position. In this image the FKP belonging to the cheeks have offset of the real face.

Figure 6: Sample images with rotated or tilted faces.

As a future work the goal is to test the detector thoroughly and get the statistical results. After that try to precise the model with small changes or try to use more methods for data augmentation to build more robust training dataset and possible the detector as well. Also, it is planned that the results of the detector should be used to estimate the rotation (in all axis) of the face.

Bibliography

[1] Namish A., Grimberge A.K., Vyas R.: *Facial Key Points Detection using Deep Convolutional Neural Network – NaimishNet*. CVPR, 2017.

[2] Masaryk University: *Fidentis – User guide*. [Online; visited 20.12.2019]. URL https://www.fidentis.cz/user-guide/

[3] Nishad G.: *Facial Keypoint Detection: Detect relevant features of face in a go using CNN & your own dataset in Python*. [Online; visited 20.12.2019]. URL https://towardsdatascience.com/facial-keypoint-detection-detect-relevant-features-of-face-in-a-go-using-cnn-your-own-dataset-e09cf359c2bc

[4] Wu W.: Wider *Facial Landmarks in-the-wild*. [Online; visited 20.12.2019]. URL https://wywu.github.io/projects/LAB/WFLW.html

[5] Wu W., Chen Q., Yang, S., Wang Q., Cai. Y., Zhou Q.: *Look at Boundary: A Boundary-Aware Face Alignment Algorithm*. CVPR, 2018.

[6] Kanich, O.; Dvořák, M.; Drahanský, M.: *Generátor a detektor modelů zbraní (Generator and detector of weapon models)*, Brno University of Technology, Faculty of Information technology, Brno, Technical report for project VI20172020068 (VRASSEO), 2019.