March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

Deep Learning based Face Recognition System using Dual Shot Face Detector and Face Landmarks

Chang-Jin Seo
Associate Professor
Department of Information Security Engineering, Sangmyung University, Cheonan, 31066, South Korea
cjseo@smu.ac.kr

Article Info
Volume 83
Page Number: 4434 - 4440
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 26 March 2020*

Abstract

Background/Objectives: Face detection and identification system have become an essential system in the image-security-system area. The present study aims to develop a robust face detection and identification system in complex scenes using Deep Learning, Dual Shot Face Detection (DSFD) algorithm, and face landmarks. Methods/Statistical analysis: The face detection process is designed by the DSFD algorithm, which has the best accuracy rate in the face detection area recently. In addition, we use the WIDER Face Dataset at the proposed system network training, which is the famous face detection training and testing DB. In addition, the identification process is composed of the face landmarks vector information using deep learning network.

Findings: Recently, state-of-the-art face detectors can be roughly classified into two-stage (R-CNN) detection and one-stage (SSD, YOLO) methods. However, one-stage face detection architecture has fascinated more attention due to its higher inference efficiency and fast system deployment. In the experiment, we can find that the detection accuracy depends on the face detection algorithms applied to it. Improvements/Applications: In the experiment result, we could find that the proposed method showed better face recognition performance compared to the conventional SSD based face recognition method.

Keywords: *Deep Learning, Face Detection, Face Recognition, DSFD, Face Identification, Face Landmarks.*

1. Introduction

In recent years, reviews of the existing researches show that fast and robust face detectors in the complex scene have selected as an essential process for face recognition, detection, and identification methods. Although the convolution neural network (CNN) based face detectors and classification systems have been broadly accepted, detecting faces with a massive degree of variability in scale, occlusion, illumination, pose, expression, and real-world applications remain hard works. Many studies have been actively conducted on the face detection and identification of faces on the complex scenes to solve these problems [1]. In addition, some deep learning systems have been improved, implementing face detection and identification methods. Until some years ago, the use of deep learning networks in face recognition applications is limited by some weaknesses, such as reduced recognition rates, slow training of massive datasets, and problems in the learning of multilayer networks. However, speedy improvement in GPGPU and the advancement of deep learning network architectures have reignited researchers' attention in algorithms over the last ten years [2].

Besides, the performance and accuracy of face detection and identification methods have

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

enhanced due to the development of robust and several deep learning algorithms and in a fast general-purpose graphics processor unit (GPGPU). As a result, the fast face detection and identification methods using ultra-high-definition (UHD) videos, advanced GPGPU, and deep learning have been continuously entered into the face recognition system markets [3]. Of particular interest are several studies researching the detection and identification of face using FHD (full high definition) and UHD in the complex scenes. The resolution of the image sensor installed in security cameras has recently improved from FHD to UHD, which has allowed the face recognition in a broad area. The present study proposes the method of face detection and identification method using the DSFD, face landmarks, and deep learning algorithm to develop in complex scenes. Now, it is employed for real applications in various fields, including robots, unmanned security system markets, driverless cars, and artificial-intelligent agents [4]. This paper introduces a face detection and identification method that uses deep learning networks trained with WIDER Face Dataset. The rest of the paper is composed as follows. Section 2 explains SSD, YOLO, and DSFD methods that are adopted in the proposed method. Section 3 states the face detection and identification method proposed in this paper. In Section 4, the experiment results of the proposed system are presented. Finally, conclusion and future work are given in Section 5.

2. Related Works

Previous state-of-the-art face detection methods can be divided into two groups. The first one is based on the Region Proposal Network (RPN) used in Faster R-CNN and uses two-stage detection method [5]. The other one is the Single Shot Detector (SSD), and You Only Look Once (YOLO) based one-stage method, which gets discarded of RPN and directly predict bounding boxes and confidence scores [6,7]. One-stage face detection architectures have attracted more attention due to its stronger performance and fast system deployment. Despite the development performed by these methods, there have still some problems with feature learning, loss design, and anchor matching. Recent other studies on face detection methods using deep learning algorithms have adopted various techniques to search for face identification, object detection system, and security control system in complex scenes. Most face detection methods consist of a large amount of data so that most methods need much computation to detect the face. The performance of the face detection algorithm affects the detection time of the face detection and identification process. Once a complex algorithm is applied to detect for face, the detection performance can improve, but real-time performance will be reduced. In contrast, once a simple algorithm is selected, real-time processing abilities can improve, but the detection performance will be decreased.

• SSD (Single Shot Multibox Detector)

Single Shot Multibox Detector (SSD) based one-stage method eliminates RPN and directly predict bounding boxes and confidence scores [Figure 1]. Recently, one-stage face detection framework has attracted more research due to higher inference efficiency. Despite the progress carried out by the above methods, some issues still exist in three respects. Feature learning and extraction parts are essential for face detectors. Feature Pyramid Network (FPN) is currently widely used in state-of-the-art face detectors for various functions. However, the FPN aggregates the hierarchical feature maps between the high and low-level output layers that do not take into account the information in the current layer, and the context relationships between anchors are ignored. Existing loss functions used to detect loss design objects include regression losses for face areas and classification losses to identify whether faces are detected. SSD is a feedforward convolution

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

neural network that generates confidence scores on a set of bounding boxes of fixed sizes for the existence of objects in that bounding boxes. The initial network layer is used up to conv4_3 in VGG-16 and then modified. SSD adds the convolution layer following VGG-16 conv4_3. As SSD goes through the convolution layer, the size of the feature map decreases and can be used to detect objects at various scales. The feature map use for object detection in SSD models is 38 ×38, 19 × 19, 10 × 10, 5 × 5, 3 × 3, 1 × 1 size [6].

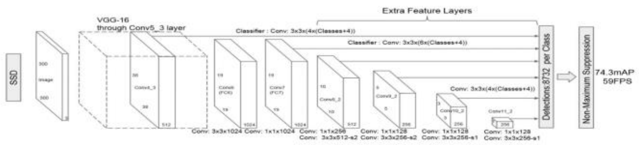

Figure 1. The architecture of Single Shot Multibox Detector [6]

• YOLO (You Only Look Once)

YOLO is one of the deep learning networks for fast object detection and has 24 convolution layers developed by 2 connected layers [Figure 2]. It uses 1×1 reduction layers and 3x3 convolution layers instead of the inception module used by GoogLeNet. Reduces computations to enable fast object detection. It applies a single neural network to the entire image. A single neural network predicts the probability of a class and bounding box directly across an entire image in one evaluation. This unified architecture makes YOLO very fast. It passes the input image over the network and predicts bounding boxes and class probabilities. YOLO divides the input image into 7x7 grids (i.e., 49 grid cells). For each grid cell, two bounding boxes of varying sizes are created to locate objects that are centered inside the grid cell. The network predicts class probabilities for each of the 98 boundary boxes and 49 grid cells per image. It calculates a confidence score each bounding box, so get 98 confidence scores. The confidence score reflects the likelihood that the box will contain objects and how accurate the bounding box is. The higher the confidence score, the thicker the rendering box. YOLO applies non-maximum suppression (NMS) to remove the less reliable bounding box [7]. This feature shows with high confidence scores as the final prediction. The YOLO algorithm is a fast object detection method, but there are some problems with small size object detection.

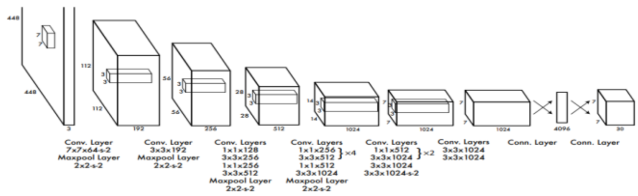

Figure 2. The architecture of You Only Look Once [7]

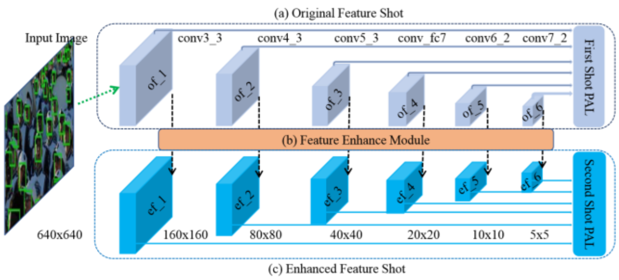

Figure 3. The architecture of Dual Shot Face Detector [8]

• DSFD (Dual Shot Face Detector)

DSFD is a dual-shot face detection method that sets the face detector to high performance using two separate anchors, even if the face size difference of the image is large. DSFD network proposed a feature enhanced module (FEM) that combines the advantages of feature pyramid network (FPN) and receive field block (RFB) to extend dual shot detection performance in a single shot and adopted two different anchor pressure (FEMs) to detect dual shots' detection performance better [8]. The new anchor allocation strategy also proposed improved anchor matching (IAM) to integrate data growth to improve face detection performance to improve regulator initialization. The architecture of the face detector DSFD is shown in [Figure 3], showing the structure using FEM (b) to obtain enhanced PALs

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

from the VGG basic feature extraction structure to produce an improved (c) from the original feature shot (a).

3. Proposed Method

This paper is designed for the face detection process through the DSFD algorithm, which is recently known as the best face detection method. We also use deep learning networks and face landmarks vector information to identify faces in complex scenes. The proposed system is designed for robust face detection trained with the CNN model, and we use the WIDER Face dataset that included clear face images, but excluding blurred, occluded, and low-light images on the training process. We also are designed a face identification method using a deep learning network using face landmarks to recognize a person's face in complex scenes.

• Implementation of the Proposed Method

The proposed method, based on a deep learning network, aims to detect and identify faces in complex scenes. The face detection and identification process are implemented using DSFD, face landmarks, and deep learning networks. DSFD improves some weaknesses in the SSD algorithm and outperforms other object detection algorithms such as Faster R-CNN and YOLO. The DSFD architecture uses the same extended VGG-16 backbone as the PyramidBox and S3FD [8]. We select conv3_3, conv4_3, conv5_3, conv_fc7, conv6_2, conv7_2 as the first layer of shot detection to create 6 feature maps called feature map1-6. Then, use the feature-enhanced module to make the second shot detection layer by entering the same size as the original feature map and an SSD-style head. The proposed method has implemented systems in SDX-4195, deep learning server for the implementation, using OpenCV 3.4, CUDA 9.1, Xeon E5-2650 4CPU, GTX-1080Ti 4-GPU, WIDER Face Dataset, torch, and torchvison.

• Network Training

In the network training, we use the WIDER Face Dataset for the training process of face detection, which has 393,703 face labels in 32,203 images. The dataset has a variety of poses, scales, occlusion, expression, makeup, and illumination in addition to a typical face [Figure 4] and is organized based on 61 event classes. For each event class, we randomly select 40%, 10%, 50% data as training, validation, and testing sets [9]. Therefore, a reasonable assessment of the general face detection model is possible. However, we use only a train and validation dataset other than the test images set and excluded ambiguous Ground Truth. The minimum face size used for learning is 10x10, but during the experiment, images of faces size less than 15x15 are excluded from the face identification process because it was too small to identify. Therefore, 97,532 faces information is used in 15,374 images for face detection. We present the following details of network implementation: Also, the pre-trained VGG was initialized as the ImageNet backbone network. All new convolution layer parameters are initialized using the 'Xavier' method. For the fine-tuning of the DSFD model, we use SGD with 0.9 momentum, 0.0005 weight decay. The batch size used for network training is set to 16, and the learning rate is set to 0.001 at the first 20k step, and then to 0.0001, 0.00001 at the next 10k step.

Figure 4. WIDER FACE Dataset [9]

• Detection of Face

In the detecting of face, we employed the feature-enhanced module to the performance of the DSFD

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

and applied it to the face detection process. First, the validation set of the WIDER FACE dataset is used to evaluate the performance of the proposed model. Second, to analyze the conventional SSD and performance, we applied the same test data and validation set to detect the face and found that the model with FEM applied to the DSFD has an excellent performance in the face detection. Third, after the performance evaluation, 427 faces are detected in Image DB to detect each face, and the face landmarks are extracted from each detected face and used as learning data for the face identification classifier. For analyzing the performance, after DSFD training, the face detection performance of the proposed model is used in the experiment data of scale, pose, occlusion, and small, medium, and large size images [Figure 5]. [Figure 6] shows the entire process used for the proposed experiment.

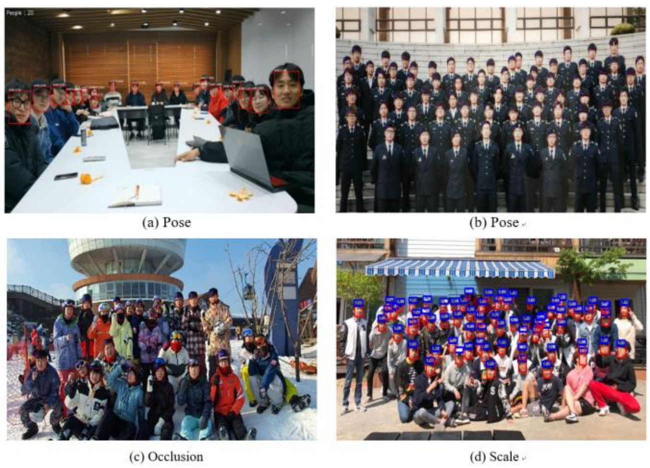

Figure 5. The face detection results with test images dataset

• Face Identification

shows 68 face landmarks in the image. Now that we know where the eyes, mouth, nose, eyebrows, and face contours are located, we now have a simple rotation of the image, scale, and shear so that the eyes and mouth can be as centered as possible. In this experiment, we did not use 3D warps. It distorts the image. We use primary image transformations, such as rotation and scale, to preserve horizontal lines. We use 68 face-landmarks information to generate 128 vector values for each face with a pre-trained deep learning network [11]. In this experiment, we configure the system to compare the already stored face image DB information with the generated vector information using the deep learning classifier to identify the final face.

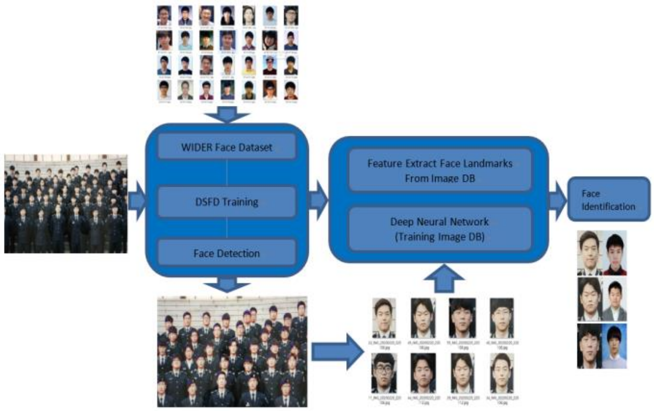

Figure 6. The Face Detection and Identification Process of the proposed method

Figure 7. The Face Landmarks of the proposed method

4. Experimental Result

In the experiment, the detection network is employed to face detection as a model network by learning DSFD, and the detected face is used to generate face landmarks information and applied to the face detection process. WIDER Face dataset is used to train the face detection network and includes Easy, Medium, and Hard labels information considering detection difficulty. These configured datasets are prepared by separating them into train, validation, and networks are trained by evaluating confidence scores of predicted bounding boxes based on the threshold 0.5 of the predicted bounding box and the IoU of Ground Truth information. The purpose of this paper is to identify the face of the image DB and detected faces. So, the test dataset is not

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

used in this experiment. The minimum face size used in the learning process was 10x10. However, when the size of the face detected in the experiment was less than 15x15, it was not satisfying for the face identification process. So, the minimum face size is used in the experiment only when the face size was 15x15 or larger. DSFD network configured CNN with VGG-16 structure as a feature extractor and networked with FEM. When assessing the performance of the Easy, Medium, and Hard set in the WIDER Face dataset, the performance of the network 93.5%, 91.2%, and 84.5%. The Image DB used in the experiment consisted of a total of 427-face information, and the input faces are divided into (small < 150 pixels, medium < 400 pixels, and large > 400 pixels) and used in the experiment using 682 test images of scale, pose and occlusion [Figure 8].

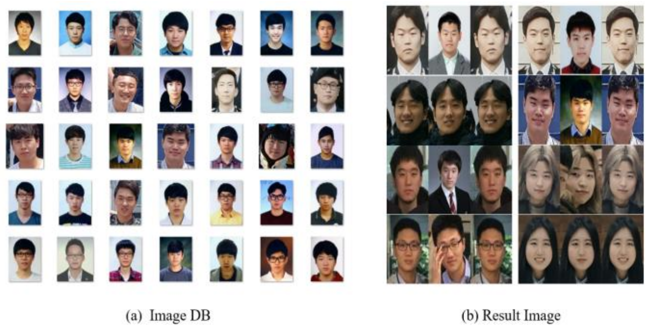

Figure 8. The Image DB(a) and Face Identification results with images dataset(b)

The experiment results showed that the performance of the DSFD is rated to be 1.5% better in small, 0.8% in medium, and 0.5% in large, compared to the performance of the SSD method and especially the size of the face is small, it is analyzed that the performance of the face detection is better. In the face recognition process, 68 face landmarks information of detected faces is converted to 128 vectors. Also, it used in the identification process with information from pre-learned Image DB. Once faces are detected, the identification performance was 94.8% in small, 98.1% in medium, and 98.3% in large size images during evaluating with Image DB [Table 1]. In the experiment, the case of face identification failure occurred in the following cases: The detected face is too small, changes the glasses, two or more faces are occluded, and the face shape is different from the DB.

Table 1. The experimental result

Test Face Size	Model	Detection	Identification
Small	SSD	87.72%	93.25%
	Proposed Method	89.62%	94.82%
Medium	SSD	94.54%	96.42%
	Proposed Method	95.47%	98.15%
Large	SSD	94.92%	96.83%
	Proposed Method	96.83%	98.33%

5. Conclusion

The present study is designed and implemented the face detection and identification system using DSFD, face landmarks, and deep learning algorithm. In the experiment result, we could find that the proposed method is more accurate than the conventional face detection system based on SSD. We use DSFD algorithm to detect faces and trained with the WIDER face dataset that showed excellent performance in face detection and used face landmarks information to identify in the face identification process. Because of the experiment, DSFD network caused some problems when the hat helmet, glasses, low-level resolution, and faces occlusion images. In the future, the proposed approach will be studied further with an extended set of images taken at the different shapes of the face, and its performance in detecting and identifying different kinds of objects will be examined.

6. Acknowledgment

This research was funded by a 2020 research Grant from Sangmyung University

TEST Engineering & Management

March - April 2020
ISSN: 0193-4120 Page No. 4434 - 4440

References

[1] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, Jia Li. Salient object detection: A survey: Computational Visual Media. 2019; 5(2):117–150. https://doi.org/10.1007/s41095-019-0149-9

[2] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature, 2015 (521):436-444. DOI: https://doi.org/10.1038/nature14539

[3] Mei Wang, Weihong Deng. Deep Face Recognition: A Survey: Computer Vision and Pattern Recognition, 2019: arXiv:1804.06655v8

[4] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition: ILCLR 2015. arXiv:1409.1556v6

[5] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks: IEEE Transactions on Pattern Analysis and Machine Intelligence. Jun2 2017 39(6):1137–1149. arXiv:1506.01497v3

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single Shot Multibox Detector: In Proceedings of European conference on computer vision (ECCV). 2015. https://doi.org/10.1007/978-3-319-46448-0_2

[7] Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali. You Only Look Once: Unified, Real-Time Object Detection: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:779-788. arXiv:1506.02640v5

[8] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, Feiyue Huang. DSFD: Dual Shot Face Detector: Computer Vision and Pattern Recognition. 2019. arXiv:1810.10220v3

[9] Shuo Yang, Ping Luo, Chen Change Loy, Xiaoou Tang. WIDER FACE: A Face Detection Benchmark: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). DOI : 10.1109/CVPR.2016.596

[10] Vahid Kazemi, Josephine Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014 June:1867–1874. https://doi.org/10.1109/CVPR.2014.241

[11] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering: in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015. DOI: 10.1109/CVPR.2015.7298682