Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/36/20230441

Face perception in deep learning and safety

Yiyun Xie

Electronic and Computer Engineering, Rutgers University, New Brunswick, USA

yx291@roseprogram.rutgers.edu

Abstract. Face perception are very useful and common in our daily life today. Not only in the unlock phone but also in some crime situation. This article will introduce several way for computer to do the face perception. In this article, I will introduce the modern majority way to do the face perception. The common way to reading face is whether separate in different part or building 3D figure for face. There are several advantage by using different way like Face net, Fece++. They could be used in different condition. After reading the image from picture, we need to train computer to do the match and do better. This is the machine learning, the best tool today is CNN or python. Even all of face perception tool could not handle the low quality picture or video, but they could have 90% accuracy in certain condition. That is the accuracy for human to do it, but computer have faster speed in face perception than people did.

Keywords: Face Perception, CNN, Machine Learning.

1. Introduction

Recent years have witnessed extensive exploration into face detection and recognition studies. The utility of face recognition applications has gained prominence across security and various other domains. Applications encompass camera surveillance, authentication on contemporary electronic devices, criminal investigations, management of database systems, and smart card utilization, among others. In light of this, the present work introduces deep face recognition learning algorithms, intended to enhance the precision of recognition and detection processes.

The central objective of facial recognition revolves around verifying and identifying facial attributes. The real-time capture of these attributes is followed by processing through haar cascade detection. This sequential workflow can be segmented into three distinct phases. The initial phase entails the detection of faces via the camera. Subsequently, the second phase encompasses an analysis of the acquired input, incorporating features and databases. These elements work in conjunction with the construction of a neural network support model. Concluding the process, the final phase involves facial authentication, resulting in the computer assigning a specific score for graphic comparison.

2. The way to read image

Today, there are several ways to let the computer do face perception. The first one is face++[1]. They collect 10000 human pictures online and train their model to do face perception. They get 99.6% accuracy to determine the same person's face at a certain time. The way they use to determine the human face is separate the whole into different parts: eyebrow, eye center, nose tip, and mouth corner. (shown in the figure 1) And then training the model to similar the graph through each part and then do

DOI: 10.54254/2755-2721/36/20230441

the determine. However, they get low accuracy on the way to determining face in the video and for the same people of different ages. They only get 60% accuracy to determine a person's face at different ages, but the real person determines the rate is 90%.

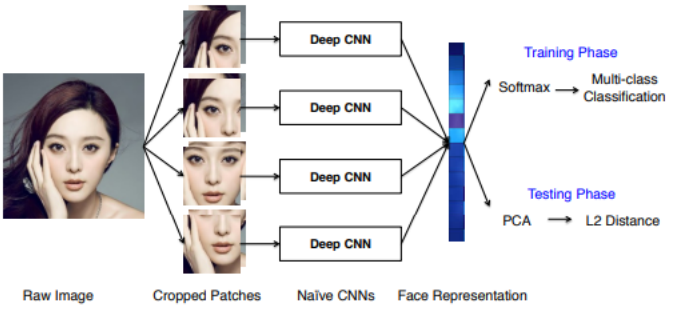


Figure 1. This figure show how face++ separate face in different part [1].

In this case, this way is not that much effect with people in different making up and ages. It only works for the human picture. However, this is a new way to do the face perception. However, in the section, it is not working so well. Author analysis some reasons their way is not working well for the different ages and videos. Most of the faces the web collects are from celebrities: smiling, made up, young, and beautiful. It is far from the images captured in everyday life. Although the accuracy of the online picture is high, its performance is still difficult to meet the requirements of real applications. Pose, occlusion, and age changes are the most common factors affecting system performance. However, we lack adequate investigations of these intersection factors and effective methods to address them clearly and comprehensively. So the author mentions their future work to improve this. For example, video is one of the data sources that can provide a large amount of spontaneous weakly labeled face data, but it has not been fully explored and applied to large-scale face recognition. On the other hand, data synthesis is another direction to generate more data. For example, manually collecting data with within-person age variation is very difficult. So a solid age change generator can be of great help. A 3D facial reconstruction is also a powerful tool for synthetic data, especially in modeling physical factors.

In this case, 3D facial reconstruction is the majority way to deal with face perception. Because there are a lot of pictures from the same people who are in different positions or taking in different angles. Those are the very important scale for determining face. To let a computer determine those we need to build a 3D model for a human face that can use in different angles and light sources.

In this article, the author describes a system, including analytical 3D modeling of Fiduciary-based faces for crop-warping detected faces into 3D frontal mode. (shown in the figure 2) which is a common way to deal with faces in video. "In order to align faces undergoing out-of-plane rotations, we use a generic 3D shape model and register a 3D affine camera, which is used to warp the 2Daligned crop to the image plane of the 3D shape" [2] In the end they got 97% accuracy to deal with the video face. This is much higher accuracy than humans did. An ideal face classifier would only be able to match human faces with accuracy. Descriptors need to be invariant to pose, lighting,Expressive power and image quality. should also be general, It can be applied to the feeling of different groups of people. There are hardly any modifications (if any). Also, short descriptors are preferable, as are sparse features if possible. Of course, fast computation time is also an issue. In this model, the time cost to deal with a picture is 0.33 which while takes a long time to deal with the face in the video.

DOI: 10.54254/2755-2721/36/20230441

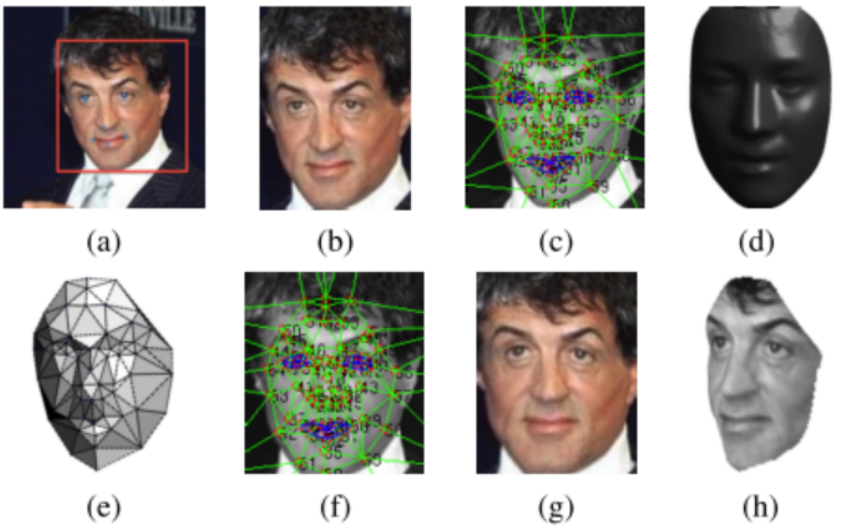


Figure 2. This figure show how to build model for face [2].

Within the context of Facial Recognition (FR) systems, a central hurdle pertains to the formulation of an effective feature representation approach. The goal is to efficiently extract relevant attributes from a given biometric source, enhancing the quality of representation. Among the steps critical to image classification, feature extraction emerges as a pivotal aspect. This process entails retaining paramount information crucial for accurate classification.

In the realm of FR systems, a primary hurdle lies in devising a feature representation strategy that effectively extracts and utilizes biometric attributes. The objective is to optimize the selection of representation for a given biometric. Among the critical stages of image classification, feature extraction stands out. This process entails safeguarding the essential information pivotal for accurate classification. Over time, a plethora of feature extraction techniques tailored to biometric systems have been introduced. Notable among these are principal component analysis (PCA) [3], independent component analysis (ICA) [4], local binary patterns (LBP) [5], and histogram-based approaches [6].

Recent emphasis has been placed on leveraging deep learning, particularly convolutional neural networks (CNNs), as the preferred feature extraction method for FR. This approach offers considerable advantages [7].

The forthcoming discussion will concentrate on the most prevalent method, Convolutional Neural Networks (CNNs). Various strategies exist for employing CNNs. Firstly, a model can be constructed from the ground up, tailoring it specifically to the task at hand. In this scenario, an architecture from a pre-existing model is utilized as a foundation, which is then fine-tuned on the target dataset. Alternatively, when dealing with extensive datasets, transfer learning can be harnessed. (shown in figure 3) This involves leveraging pre-trained CNN features, yielding notable advantages. Lastly, transfer learning via CNNs can be executed by maintaining the convolutional base in its original configuration and channeling its output to the subsequent classifier.[8]

DOI: 10.54254/2755-2721/36/20230441

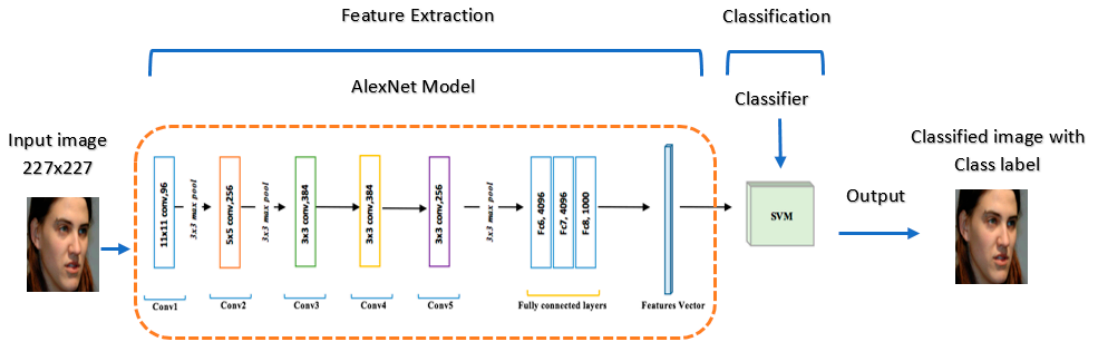


Figure 3. This figure shows how to handle the face changed throw the years [8].

Ultimately, the researchers utilize a Convolutional Neural Network (CNN) to train facial data extracted from sources such as YouTube. The attained accuracy reaches approximately 94%, aligning with anticipated outcomes. The author's future plans encompass expanding their dataset by incorporating a larger array of facial images and videos for enhanced model training. By doing so, they aim to bolster their model's capabilities.

To validate their model's effectiveness, a comparative analysis is conducted against cutting-edge counterparts across various datasets including FEI Faces, LFW, YouTube Faces, and ORL. Impressively, the results underscore that their model surpasses the accuracy achieved by most state-of-the-art models. Looking ahead, the researchers remain committed to advancing both recognition and classification accuracy. This endeavor entails the inclusion of more diverse databases for CNN model refinement, along with the exploration of varied convolutional neural network models to optimize functionality. The augmentation of Facial Recognition (FR) models introduces potential avenues for innovative feature extraction techniques.

3. The machine learning

Biometric software plays a growing role in contemporary security, administrative, and corporate systems. Biometric methods encompass fingerprints, retinal scans, voice recognition, and facial detection. Facial identification is particularly intriguing for discreet, unobtrusive detection, and verification without explicit consent. This technology finds application in areas like airport security and gains traction in law enforcement due to ongoing system enhancements and face database expansion. In this instance, the author employs a model to contrast facial features during different life stages, potentially valuable for locating lost children. The comparison of facial vectors relies on matching image attributes, facilitated by the Python library for template-based facial analysis. It processes facial attributes from suspicious input images and cross-references them with criminal records to generate and juxtapose identifying features.[9]

Python Facepplib

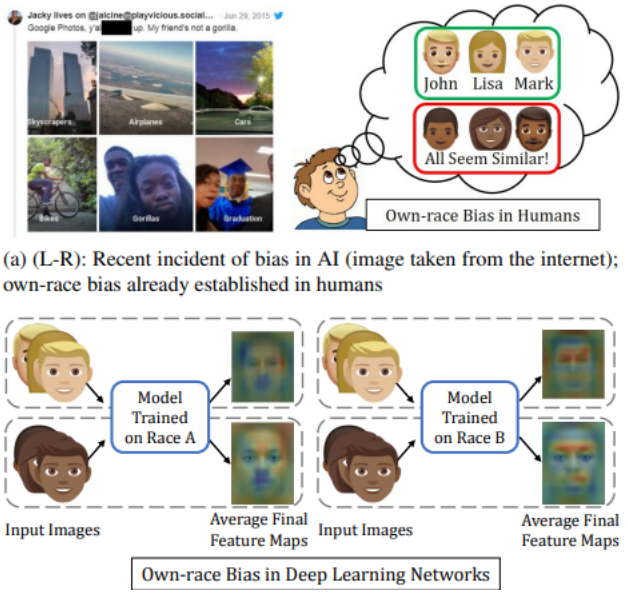
Python serves as a versatile and well-documented programming language applicable to diverse scenarios. In the realm of Computer Vision, Python stands out due to its user-friendly nature. The Facepplib API offers a straightforward template matching process. This configuration entails setting up the api_key and Secret_key.

The API supplies a comparative technique that reveals resemblances between two templates. Confidence levels emerge from gauging the disparity between attributes. Proximity indicates similarity; if characteristics are close, their likeness is probable. Each facial set shares a confidence level for certain attributes. Variations in analysis dictate different thresholds for facial comparisons. When confidence surpasses the threshold, the faces are labeled as matching, potentially leading to the identification of criminals or missing persons. Pertinent details about web-based facial images

DOI: 10.54254/2755-2721/36/20230441

accompany this process. Ultimately, the author reports a 90% accuracy through online picture comparisons, a tool holding significant utility, especially in missing children cases.

When age problems and low-quality problems could be solved. Shruti Nagpal found out that race bias is part of the problem that influences accuracy. Based on the criminal rate of African Americans is higher than local people. Some FR programs will rate African American higher criminal rates than local ones.[10] In this case, the authors do the research to improve this part. This study goes beyond the existing literature Analyzing Deep Learning-Based Face Recognition Models. (shown in figure 4)



(a) (L-R): Recent incident of bias in AI (image taken from the internet); own-race bias already established in humans

Figure 4. This figure shown how traditional AI read human face [10].

Bias is present in the study's findings. For the initial investigation, this study aims to comprehend the nature of deep network encoding and its resemblance to human cognitive functions in the context of face recognition, while recognizing distinct parameters. We examined four deep learning networks, whether derived from spontaneous creation or pre-trained on comprehensive datasets containing a significant volume of around 10 million images. Our intention was to deduce our conclusions from this extensive evaluation.

We hold the belief that unraveling the mechanisms behind the operation of deep networks can contribute to the development of more equitable AI systems, capable of impartial decision-making. Since deep learning models exhibit similarities to human cognitive processes, the outcomes of this study may offer the potential to enhance the utilization of pre-existing systems. This can be achieved through the implementation of appropriate techniques and cognitive methodologies to counteract biases.

A novel endeavor, this study strives to determine if biases are embedded within face recognition systems and their specific occurrences. The study encompasses two distinct case studies involving race and age. The analysis of numerous deep-learning networks reveals that biases related to race and age are indeed present within facial recognition models. Analogous to human behavior, deep learning networks also display discernible inclinations.

These networks tend to concentrate heavily on distinct facial regions associated with specific ethnicities, although discrepancies exist among different racial groups. To the best of our knowledge, this study represents the foremost instance that demonstrates these partialities in behavior.

Deep learning networks exhibit parallels to human cognitive processes, thereby offering avenues for beneficial applications aimed at rectifying biases within these networks. Our analysis demonstrates the advantages of employing extensive datasets for training purposes. However, it's important to note

DOI: 10.54254/2755-2721/36/20230441

that relying solely on large-scale data isn't recommended as a comprehensive remedy for eliminating biases. Instead, careful attention should be directed toward curating training data for deep learning models, ensuring incorporation of a wide spectrum of variability.

Similar tendencies emerged from a cross-age study, wherein facial recognition processes unconsciously incorporate age information relative to training samples. This results in distinct regions of interest for different demographic groups, thereby introducing ownership bias. Even though modern facial recognition systems exhibit state-of-the-art performance, they still exhibit a notable degree of age bias, particularly when applied to children's facial images.

Among the authors, apart from CNN, one author employs an alternative approach to system development. This method involves inputting the face under scrutiny into the model to obtain a numerical degree for verification.[11] Deep network-based face recognition methods utilize classification layers trained on face identity data. The intermediate bottleneck layer subsequently serves as a generalized recognition representation. Notably, this method's drawback is its indirectness and inefficiency. It relies on the assumption that the bottleneck representation can effectively generalize new faces. The use of bottleneck layers often results in substantial representation size per face, mitigated to some extent by recent work involving PCA. Nevertheless, such transformations are linear and can be learned through a single network layer. In contrast, FaceNet directly trains by generating a compact 128-dimensional embedding using a triplet loss function. This triplet comprises two matching face thumbnails and one non-matching face thumbnail. Positive and negative values are separated by a specific distance through the use of face thumbnails and associated losses. The thumbnails are cropped tightly from facial regions, devoid of 2D or 3D alignment, except for zooming and panning.

Evaluations indicate that incremental training yields superior performance in terms of both speed and dataset size compared to batch training. Interestingly, reduced training speed and dataset size do not adversely affect accuracy; conversely, the employed face detection method can significantly enhance face recognition accuracy. In the realm of face detection, deep learning methods like MTCNN and YOLO-face outperform traditional methods such as Viola Jones and LBP. Consequently, deep learning methods present themselves as more suitable for face recognition tasks.

Comparison between MTCNN and YOLO-face reveals slight discrepancies in evaluation results, largely contingent on the dataset used. YOLO-face demonstrates superiority in detection accuracy and speed. This method's enhanced performance renders it an excellent candidate for the development of accurate proctoring systems. Such advancements in facial recognition accuracy and memory utilization will inevitably contribute to enhanced user verification and seamless integration within devices, particularly when applied to mobile online exam invigilation systems. This multifaceted progress lays the foundation for comprehensive analysis and further exploration.

Other novel methods or algorithms in the field of face detection. This a good try compared with the traditional way.

On the hand, face perception is also used to find criminal [11]. Face embedding need to be created so that they can be compared with different vectors. Here are the steps to create embedding using the FaceNet model. After loading the compressed file of detected faces, the pixel values need to be normalized according to Face Net requirements. Load a preordained Keras Face net model. Enumerate each face to find its predictions and embedding from the training and test sets. Embedding are saved as compressed Bumpy arrays.

In this part of the process, a machine learning model is used to classify the embeddings to identify them as one of the criminals. Model vector normalization is applied to scale the values before classification is applied. The sci kit-learn normalization library is used for this purpose. Next, the criminal's name is converted from string format to integer format. This is done using sci kit learn's LabelEncoder. The classification model used is a linear support vector machine because it can effectively distinguish face embedding. A linear SVM model is fitted to the training data.

DOI: 10.54254/2755-2721/36/20230441

To visualize the working of the entire model, a face was picked from the compressed test set. Then create an embed of that image. This face embedding is used as input to fit the model and obtain predictions.

4. Conclusion

Within this article, the authors propose a novel strategy geared towards addressing criminal identification, thus bolstering social security measures.

The first approach centers around face++, and its core revolves around the accumulation of extensive datasets. The model is educated to discern facial variations through increasingly comprehensive data sets. This method finds suitability within large corporations such as Baidu, Tencent, and Google. The potential for amassing progressively comprehensive training data sets looms promising. This inclusive dataset encompasses images of individuals across distinct age brackets, diverse ethnicities, and various aesthetics (ranging from aesthetically pleasing to less so). The incorporation of such diverse data cultivates the model's adeptness in accommodating disparities present in real-world applications.

The second method, termed FR+FCN, employs a synthesis-oriented strategy. Utilizing 3D models, this approach amalgamates assorted facial types, thereby augmenting the dataset. This technique boasts economic efficiency, rendering it an apt choice for widespread implementation. Notably, CNN-3DMM estimation is a standout component of this approach, as the author provides comprehensive source code to facilitate further exploration and reference.

Under optimal conditions, both of the aforementioned methodologies yield face recognition accuracy that equals or surpasses human capabilities. Nevertheless, the practical utilization of face recognition is hampered by variances such as lighting, angles, expressions, and age. These elements underscore the challenge in deploying face recognition technology on a broad scale. Moving forward, regardless of the approach adopted, the ultimate objective revolves around enhancing the model's adaptability within complex real-world environments. It is imperative that the technology attains human-level accuracy even in intricate scenarios.

Face recognition technology, having attained maturity in the domain of computer vision and deep learning, stands poised for widespread integration. The anticipation is palpable, as we await the extensive application of this transformative technology.

References

[1] E. Zhou, Z. Cao, and Q. Yin, 2015, *arXiv,* 19.
[2] S. Almabdy and L. Elrefaei, 2019, Applied *Sciences*, **9**, 20.
[3] Kshirsagar, V.P.; Baviskar, M.R.; Gaikwad, M.E. 2011 *In Proceedings of the 2011 3rd International Conference on Computer Research and Development*, Shanghai, China, **2,** 11.
[4] Bartlett, M.S.; Movellan, J.R.; Sejnowski, T.J. 2002, *Neural Netw*. **13**, 1450
[5] Ojala, T.; Pietikainen, M.; Maenpaa, T. 2002, *IEEE Trans*. Pattern Anal. Mach. Intell. **24**, 971
[6] Liu, Y.; Lin, M.; Huang, W.; Liang, J. 2017, *J. Vis*. Lang. Comput. **43**, 103
[7] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. 2014, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* **23**, 1.
[8] S.Almabdy and L. Elrefaei, 2019 *Applied Sciences*, **9**, 20.
[9] S. Ayyappan and S. Matilda, 2020, *Automation and Networking (ICSCAN)*, Pondicherry, India: IEEE.
[10] S. Nagpal, M. Singh, R. Singh, and M. Vatsa, 2016, *arXiv,* **19**.
[11] F. Schroff, D. Kalenichenko, and J. Philbin, 2015, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.