Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

A Review of Three Dimensional Face Recognition Algorithm Using Deep Learning Methods

M. Sivajothi [1*], V. Prema [2] and Grasha Jacob [3]

[1]Associate Professor, Department of Computer Science, Sri Parasakthi College for Women, Courtallam, TamilNadu, India E-mail: rksmano@gmail.com

[2]Research Scholar, Rani Anna Government College for Women,Gandhinagar, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, TamilNadu, India E-mail: premarishi77@gmail.com

[3] Associate Professor & Head , Rani Anna Government College for Women,Gandhinagar, Manonmaniam Sundaranar University ,Abishekapatti, Tirunelveli-627012,TamilNadu, India E-mail: grasharanjit@gmail.com

ABSTRACT

Deep learning applies multiple processing layers to learn representations of data with multiple levels of feature extraction. This emerging technique has reshaped the research landscape of face recognition (FR) since 2014, launched by the breakthroughs of Deepface method. Since then, deep FR technique, which leverages hierarchical architecture to stitch together pixels into invariant face representation, has dramatically improved the state-of-the-art performance and fostered successful real-world applications. In this survey, we provide a comprehensive review of the recent developments on deep FR, covering both broad topics on algorithm designs, databases and protocols, and application scenes. First, we summarize different network architectures and loss functions proposed in the rapid evolution of the deep FR methods. Second, the related face processing methods are categorized into two classes: “one-to-many augmentation” and “many-to-one normalization”. Then, we summarize and compare the commonly used databases for both model training and evaluation.

KEYWORDS: Deep learning, Face recognition, CNN, GAN model.

***Corresponding author:**

Dr. M. Sivajothi

Associate Professor,

Department of Computer Science,

Sri Parasakthi College for Women,

Courtallam, Tamil Nadu, India

E-mail: rksmano@gmail.com

I.INTRODUCTION

Face recognition (FR) has been the prominent biometric technique for identity authentication and has been widely used in many areas, such as military, finance, public security and daily life. FR has been a long-standing research topic in the CVPR community. In the early 1990s, the study of FR became popular following the introduction of the historical Eigenface approach. The milestones of feature-based FR over the past years are presented in Fig. 1, in which the times of four major technical streams are highlighted. The holistic approaches derive the low-dimensional representation through certain distribution assumptions, such as linear subspace, manifold and sparse representation. This idea dominated the FR community in the 1990s and 2000s. However, a well-known problem is that these theoretically plausible holistic methods fail to address the uncontrolled facial changes that deviate from their prior assumptions. In the early 2000s, this problem gave rise to local-feature-based FR. Gabor and LBP[6] as well as their multilevel and high-dimensional extensions achieved robust performance through some in-variant properties of local filtering. Unfortunately, handcrafted features suffered from a lack of distinctiveness and compactness. In the early 2010s, learning-based local descriptors were introduced to the FR community[23, 24], in which local filters are learned for better distinctiveness, and the encoding codebook is learned for better compactness. However, these shallow representations still have an inevitable limitation on robustness against the complex nonlinear facial appearance variations.

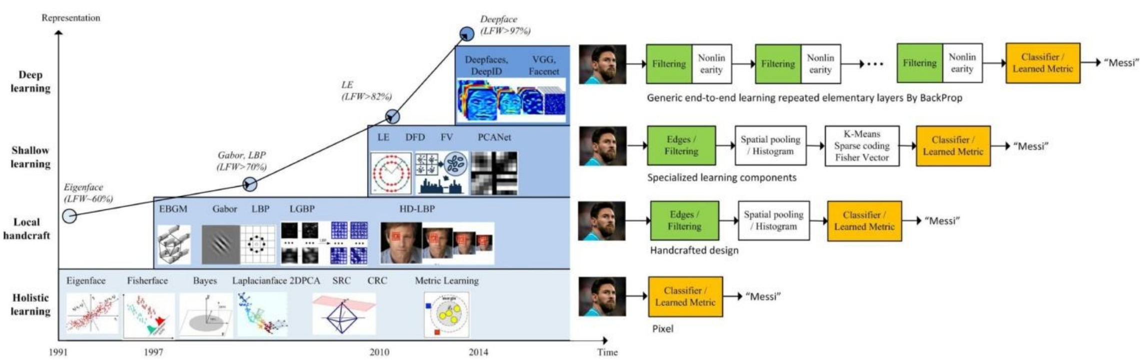

FIG. 1. Milestones Of Face Representation For Recognition.

The holistic approaches dominated the face recognition community in the 1990s. In the early 2000s, handcrafted local descriptors became popular, and the local feature learning approach was introduced in the late 2000s. Deep learning technique has reshaped the research landscape of FR in almost all aspects such as algorithm designs, training/test data sets, application scenarios and even the evaluation protocols. Therefore, it is of great significance to review the breakthrough and rapid development process in recent years. There have been several surveys on FR [20], [3]and its sub domains, and they mostly summarized and compared a diverse set of techniques related to a specific

FR scene, such as illumination-invariant FR 3D FR pose invariant FR .Unfortunately, due to their earlier publication dates, none of them covered the deep learning methodology that is most successful nowadays. This survey focuses only on recognition problem.

II.OVERVIEW

A. Background Concepts and Terminology

FR can be categorized as face verification and face identification. In either scenario, a set of known subjects is initially enrolled in the system (the gallery), and during testing, a new subject (the probe) is presented. Face verification computes one-to-one similarity between the gallery and probe to determine whether the two images are of the same subject, whereas face identification computes one-to-many similarity to determine the specific identity of a probe face. When the probe appears in the gallery identities, this is referred to as closed-set identification; when the probes include those who are not in the gallery; this is open set identification.

B. Components of Face Recognition

Before a face image is fed to an FR module, face anti-spoofing, which recognizes whether the face is live or spoofed, can avoid different types of attacks. FR module consists of face processing, deep feature extraction and face matching, and it can be described as follows:

$$M[F\ (P_i(I_i));\ F\ (P_j(I_j))] \quad (1)$$

where I_i and I_j are two face images, respectively; P stands for face processing to handle intra-personal variations, such as poses, illuminations, expressions and occlusions; F denotes feature extraction, which encodes the identity information; and M means a face matching algorithm used to compute similarity scores.

1) Face Processing: Although deep learning-based approaches have been widely used due to their powerful representation, proved that various conditions, such as poses, illuminations, expressions and occlusions, still affect the performance of deep FR and that face processing is beneficial, particularly for poses. Since pose variation is widely regarded as a major challenge in automatic FR applications, we mainly summarize the deep methods of face processing for poses in this paper. Other variations can be solved by similar methods.

“One-to-many augmentation”: generating many patches or images of the pose variability from a single image to enable deep networks to learn pose-invariant representations.

“Many-to-one normalization”: recovering the canonical view of face images from one or many images of a nonfrontal view; then, FR can be performed as if it were under controlled conditions.

2) Deep Feature Extraction: Network Architecture. The architectures can be categorized as backbone and assembled networks,. Inspired by the extraordinary success on the ImageNet challenge, the typical CNN architectures, such as AlexNet, VGGNet, GoogleNet, ResNet and SENet are introduced and widely used as the baseline model in FR (directly or slightly modified). In addition to the mainstream, there are still some novel architecture designed for FR to improve efficiency. Moreover, when adopting backbone networks as basic blocks, FR methods often train assembled networks with multiple inputs or multiple tasks. One network is for one type of input or one type of task. It provides an increase in performance after accumulating the results of assembled networks.

3) Face Matching by Deep Features: After the deep networks are trained with massive data and an appropriate loss function, each of the test images is passed through the networks to obtain a deep feature representation. Once the deep features are extracted, most methods directly calculate the similarity between two features using cosine distance or L2 distance; then, the nearest neighbor (NN) and threshold comparison are used for both identification and verification tasks. In addition to these, other methods are introduced to post process the deep features and perform the face matching efficiently and accurately, such as metric learning, sparse representation based classifier (SRC), and so forth.

III. NETWORK ARCHITECTURE

As there are billions of human faces in the earth, real-world FR can be regarded as an extremely fine-grained object classification task. For most applications, it is difficult to include the candidate faces during the training stage, which makes FR become a “zero-shot” learning task. Fortunately, since all human faces share a similar shape and texture, the representation learned from a small proportion of faces can generalize well to the rest. A straightforward way is to include as many IDs as possible in the training set. For example, Internet giants such as Facebook and Google have reported their deep FR system trained by 10^6 10^7 IDs. Unfortunately, these personal datasets, as well as prerequisite GPU clusters for distributed model training, are not accessible for academic community. Currently, public available training databases for academic research consist of only 10^3 10^5 IDs. Instead, academic community make effort to design effective loss functions and adopt deeper architectures to make deep features more discriminative using the relatively small training data sets.

1. Face Matching By Deep Features

During testing, the cosine distance and L2 distance are generally employed to measure the similarity between the deep features x_1 and x_2; then, threshold comparison and the nearest neighbor

(NN) classifier are used to make decision for verification and identification. In addition to these common methods, there are some other explorations.

2. Face Verification

Metric learning, which aims to find a new metric to make two classes more separable, can also be used for face matching based on extracted deep features. The JB model is a well-known metric learning method and proved that it can improve the performance greatly. In the JB model, a face feature x is modeled as x = + ", where and "are identity and intra-personal variations, respectively.

3. Face Identification

After cosine distance was computed, proposed a heuristic voting strategy at the similarity score level for robust multi-view combination of multiple CNN models. In extracted the local adaptive convolution features from the local regions of the face image and used the extended SRC for FR with a single sample per person. Deep features and the SVM classifier to recognize all the classes. Based on deep features, first used product quantization (PQ) to directly retrieve the top-k most similar faces and re-ranked these faces by combining similarities from deep features and the COTS matcher.

IV. FACE PROCESSING FOR TRAINING AND RECOGNITION

A. One-To-Many Augmentations

Collecting a large database is extremely expensive and time consuming. The methods of "one-to-many augmentation" can mitigate the challenges of data collection, and they can be used to augment not only training data but also the gallery of test data. we categorized them into four classes: data augmentation, 3D model, CNN model and GAN model.

Data augmentation: Common data augmentation methods consist of photometric transformations, geometric transformations, such as oversampling (multiple patches obtained by cropping at different scales) mirroring and rotating the images. Recently, data augmentation has been widely used in deep FR algorithms cropped 400 face patches varying in positions, scales, and color channels and mirrored the images. In seven CNNs with the same structure were used on seven overlapped image patches centered at different landmarks on the face region.

3D model: 3D face reconstruction is also a way to enrich the diversity of training data. There is a large number of papers about this domain, but we only focus on the 3D face reconstruction using deep methods or used for deep FR. In generated face images with new intra-class facial appearance variations, including pose, shape and expression, and then trained a 19-layer VGGNet with both real and augmented data. used generic 3D faces and rendered fixed views to reduce much of the computational effort. Employed an iterative 3D CNN by using a secondary input channel to represent the previous network's output as an image for reconstructing a 3D face. Used a multi-task CNN to divide 3D face reconstruction into neutral 3D reconstruction and expressive 3D reconstruction. Directly regressed 3D morphable face model (3DMM)[18] parameters from an input photo by a very deep CNN architecture. Synthesized face images with various poses and expressions using the 3DMM method, then reduced the gap between synthesized data and real data with the help of MMD.

CNN model: Rather than reconstructing 3D models from a 2D image and projecting it back into 2D images of different poses, CNN models can generate 2D images directly. In the multi-view perceptron (MVP) the deterministic hidden neurons learn the identity features, while the random hidden neurons capture the view features. By sampling different random neurons, the face images of different poses are synthesized. Similar to used 7 Recon codes to rotate faces into 7 different poses, and proposed a novel type of unpair supervised way to learn the face variation representation instead of supervising by Recon codes.

GAN model: After using a 3D model to generate profile face images, DA-GAN refined the images by a GAN, which combines prior knowledge of the data distribution and knowledge of faces. Combined a variation auto encoder with a GAN for augmenting data, and took advantages of both statistic and pair wise feature matching to make the training process converge faster and more stably. In addition to synthesizing diverse faces from noise, some papers also explore to disentangle the identity and variation, and synthesize new faces by exchanging them between different people. In CG-GAN a generator directly resolves each representation of input image into a variation code and a identity code and regroups these codes for cross-generating, while a discriminator ensures the reality of generated images. Identity representation of one input image and attribute representation of any other input face image, then synthesized new faces from recombining these representations. This work shows superior performance in generating realistic and identity preserving face images, even for identities outside the training dataset. Unlike previous methods that treat classifier as a spectator, A three-player GAN where the classifier cooperates together with the discriminator to compete with the generator from two different aspects, i.e. facial identity and image quality respectively.

B. Many-To-One Normalization

In contrast to "one-to-many augmentation", the methods of "many-to-one normalization" produce frontal faces and reduce appearance variability of test data to make faces align and compare easily. It can be categorized as SAE, CNN and GAN models.

The proposed stacked progressive auto encoders (SPAE) progressively map the nonfrontal face to the frontal face through a stack of several auto encoders. In a novel recurrent convolutional encoder-decoder network combined with shared identity units and recurrent pose units can render rotated objects instructed by control signals at each time step. A sparse many-to-one encoder by setting frontal face and multiple random faces as the target values.

Face identity-preserving features to reconstruct face images in the canonical view using a CNN that consists of a feature extraction module and a frontal face reconstruction module. Canonical view images according to the face images' symmetry and sharpness and then adopted a CNN to recover the frontal view images by minimizing the reconstruction loss error. A multi-task network that can rotate an arbitrary pose and illumination image to the target-pose face image by utilizing the user's remote code. Transformed nonfrontal face images to frontal images according to the displacement field of the pixels between them.

GAN model proposed a two-pathway generative adversarial network (TP-GAN) that contains four landmark-located patch networks and a global encoder-decoder network. Through combining adversarial loss, symmetry loss and identity-preserving loss, TP-GAN generates a frontal view and simultaneously preserves global structures and local details. In a disentangled representation learning generative adversarial network (DR-GAN) an encoder produces an identity representation, and a decoder synthesizes a face at the specified pose using this representation and a pose code. Incorporated 3DMM into the GAN structure to provide shape and appearance priors to guide the generator to formalization.

V. TECHNICAL CHALLENGES

In this paper, we provide a comprehensive survey of deep FR from two aspects of data and algorithms. For algorithms, some mainstream and special network architectures are presented. For data, we summarize some commonly used FR datasets. Moreover, the methods of face processing are introduced and categorized as "one-to-many augmentation" and "many-to-one normalization". Taking advantage of big annotated data, deep learning and GPUs, deep FR has dramatically improved the state-of-the-art performance and fostered successful real-world applications. As the practical and commercial use of this technology, many ideal assumptions of academic research were broken, and more and more real-world issues are emerging.

REFERENCES

1. Fg-net aging database. http://www.fgnet.rsunit.com.
2. Ms-celeb-1m challenge 3. http://trillionpairs.deepglint.com.

3. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. Pattern recognition letters, 2007; 28(14):1885–1906.
4. J. W. Abdalmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, Choi, J. Lekust, J. Kim, and P. Natarajan. Face recognition using deep multi-pose representations. In WACV, 2016; 1–9.
5. Agarwal, R. Singh, M. Vatsa, and N. Ratha. Are imageagnostic universal adversarial perturbations for face recognition difficult to detect. IEEE BTAS, 2018.
6. T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Machine Intell., 2006; 28(12):2037–2041.
7. G. Antipov, M. Baccouche, and J.-L. Dugelay. Boosting cross-age face verification via generative age normalization. In IJCB, 2017.
8. G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. arXiv preprint arXiv:1702.01983, 2017.
9. Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu. Face anti-spoofing using patch and depth-based cnns. In IJCB,. IEEE, 2017; 319–328.
10. Bansal, C. Castillo, R. Ranjan, and R. Chellappa. The dos and donts for cnn-based face verification. arXiv preprint arXiv:1705.07426, 5, 2017.
11. Bansal, A. Nanduri, C. Castillo, R. Ranjan, and R. Chellappa. Umdfaces: An annotated face dataset for training deep networks. arXiv preprint arXiv:1611.01484, 2016.
12. J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Cvae-gan: fine-grained image generation through asymmetric training. arXiv preprint arXiv:1703.10155, 2017.
13. J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Towards open-set identity preserving face synthesis. In CVPR, 2018; 6713–6722.
14. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell., 1997; 19(7):711–720.
15. J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, et al. The challenge of face recognition from digital point-and-shoot cameras. In BTAS, IEEE, 2013; 1–8.
16. S. Bianco. Large age-gap face verification by feature injection in deep networks. Pattern Recognition Letters, 2017; 90:36–42.

17. D. M. Blackburn, M. Bone, and P. J. Phillips. Face recognition vendor test 2000: evaluation report. Technical report, DEFENSE ADVANCED RESEARCH PROJECTS AGENCY ARLINGTON VA, 2001.
18. V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. IEEE Transactions on pattern analysis and machine intelligence, 2003; 25(9):1063–1074.
19. N. Bodla, J. Zheng, H. Xu, J.-C. Chen, C. Castillo, and R. Chellappa. Deep heterogeneous feature fusion for template-based face recognition. In WACV,. IEEE, 2017; 586–595.
20. K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. Computer vision and image understanding, 2006; 101(1):1–15.
21. K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy. Pose-robust face recognition via deep residual equivariant mapping. arXiv preprint arXiv:1803.00839, 2018.
22. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. arXiv preprint arXiv:1710.08092, 2017.
23. Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In CVPR, IEEE, 2010; 2707–2714.
24. T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma. Pcanet: A simple deep learning baseline for image classification? IEEE Transactions on Image Processing, 2015; 24(12):5017–5032.
25. Chen, W. Deng, and J. Du. Noisy softmax: improving the gener-alization ability of dcnn via postponing the early softmax saturation. arXiv preprint arXiv:1708.03769, 2017.