# Deep Learning models for Video based Facial Recognition Systems: A Survey

K.Sunitha

*Assistant Professor, Department of Computer Science & Engineering, Mahatma Gandhi Institute of Technology, Gandipet,Hyderabad-500075., A.P., Affiliated to JNTUH.*

***Abstract***

*Deep learning has recently achieved very promising results in a wide range of areas such as computer vision, speech recognition and natural language processing. It aims to learn hierarchical representations of data by using deep architecture models.*

*Face recognition (FR) systems for video surveillance (VS) applications attempt to accurately detect the presence of target individuals over a distributed network of cameras. Specifically, in still-to-video FR application, a single high-quality reference still image captured with still camera under controlled conditions is employed to generate a facial model to be matched later against lower-quality faces captured with video cameras under uncontrolled conditions. Current video-based FR systems can perform well on controlled scenarios, while their performance is not satisfactory in uncontrolled scenarios mainly because of the differences between the source (enrollment) and the target (operational) domains. Most of the efforts in this area have been toward the design of robust video-based FR systems in unconstrained surveillance environments. deep learning architectures proposed in the literature based on triplet-loss function (e.g., cross-correlation matching CNN, trunk-branch ensemble CNN and HaarNet) and supervised autoencoders (e.g., canonical face representation CNN) are studied.*

**Keywords:** *Deep Learning, Face Recognition, Video Surveillance, CNN.*

## I. INTRODUCTION

Face recognition (FR) systems in video surveillance (VS) has received a significant attention during the past few years. Due to the fact that the number of surveillance cameras installed in public places is increasing, it is important to build robust video-based FR systems [1]. In VS, capture conditions typically range from semi-controlled with one person in the scene (e.g. passport inspection lanes and portals at airports), to uncontrolled free-flow in cluttered scenes (e.g. airport baggage claim areas, and subway stations).

Two common types of applications in VS are:(1) Still-to-Video FR (e.g., watch-list screening), and (2) Video-to-Video FR (e.g., face re-identification or search and retrieval) [2,3,4]. In the former application, reference face images or stills of target individuals of interest are used to design facial models, while in the latter, facial models are designed using faces captured in reference videos.

The number of target references is one or very few in still-to-video FR applications, and the characteristics of the still camera(s) used for design significantly differ from the video cameras used during operations [5]. Thus, there are significant differences between the appearances of still ROI(s) and ROIs captured with surveillance cameras, according to various changes in ambient lighting, pose, blur, and occlusion [6,7]. During enrollment of target individuals, facial regions of interests (ROIs) isolated in reference still images are used to design facial models,while during operations, the ROIs of faces captured in videos are matched against these facial models. In VS, a person in a scene may be tracked along several frames, and matching scores may be accumulated over a facial trajectory (a group of ROIs that correspond to the same high-quality track of an individual) for robust spatiotemporal FR [8].

In generally, the methods proposed for still-to-video FR can be broadly categorized into two main streams: (1) conventional, and (2) deep learning methods. The conventional methods rely on hand-crafted feature extraction techniques and a pre-trained classifier along with fusion, while deep learning methods automatically learn features and classifiers conjointly using massive amounts of data. In spite of improvements achieved using the conventional methods, yet they are less robust to real-world still-to-video FR scenario. On the other hand, there exists no feature extraction technique that can overcome all the challenges encountered in VS individually[2, 9,10].

Conventional methods proposed for still-to-video FR are typically modeled as individual-specific face detectors using one- or 2-class classifiers in order to enable the system to add or remove other individuals and easily adapt over time [11,4]. Modular systems designed using individual-specific ensembles have been successfully applied in VS [3,4]. Thus, ensemble-based

methods have been shown as a reliable solution to deal with imbalanced data, where multiple face representations can be encoded into ensembles of classifiers to improve the robustness of still-to video FR [2]. Although it is challenging to design robust facial models using a single training sample, several approaches have addressed this problem, such as multiple face representations, synthetic generation of virtual faces, and using auxiliary data from other people to enlarge the training set [11, 12, 13,14]. These techniques seek to enhance the robustness of face models to intra-class variations. In multiple representations, different patches and face descriptors are employed [11,2], while 2D morphing or 3D reconstructions are used to synthesize artificial face images [12,15].

Recently, several deep learning based solutions have been proposed to learn effective face representations directly from training data through convolutional neural networks (CNNs) and nonlinear feature mappings [17, 18,19,20,21]. In such methods, different loss functions can be considered in the training process to enhance the inter-personal variations, and simultaneously reduce the intra-personal variations. They can learn non-linear and discriminative feature representations to cover the existing gaps compared to the human visual system [10], while they are computationally costly and typically require a large number of labeled data to train. To address the SSPP problem in FR, a triplet-based loss function have been introduced in [22, 23, 24, 25, 19] to discriminate between a pair of matching ROIs and a pair of non-matching ROIs. Ensemble of CNNs, such as trunk-branch ensemble CNN (TBE-CNN) [22] and HaarNet [24] have been shown to extracts features from the global appearance of faces (holistic representation), as well as, to embed asymmetrical features (local facial feature-based representations) to handle partial occlusion. Moreover, supervised autoencoders have been proposed to enforce faces with variations to be mapped to the canonical face (a well-illuminated frontal face with neutral expression) of the person in the SSPP scenario to generate robust feature representations [26,27].

## II. FACE RECOGNITION

Face recognition consists of two main tasks:

***A. Face Detection:*** where the input image is searched to find any face, then image processing cleans up the facial image for easier recognition.

***B. Face Recognition :*** where the detected and processed face is compared to the database of known faces to decide who that person is.

The difference between face detection and recognition is that in detection we just need to determine if there is some face in the image, but in recognition we want to determine whose face it is. Features extracted from a face are processed and compared with similarly processed faces present in the database.

In general, face recognition techniques can be divided into two groups:

***A. Face representation techniques :*** these techniques use holistic texture features and are applied to either whole-face or specific regions in a face image.

***B. Feature-based techniques :*** these techniques use geometric facial features (mouth, eyes, brows,

etc.), and geometric relationships between them.

Recently, many deep learning based algorithms have achieved very promising results in these two face recognition tasks.

## III. VIDEO-BASED FR THROUGH DEEP LEARNING

In video-based FR systems, facial models of target individuals are designed a priori during enrollment using a limited number of reference still images or video data. These facial models are not typically representative of faces being observed during operations due to large variations in illumination, pose, scale, occlusion, blur, and to camera inter-operability.

In contrast with shallow learning algorithms ,deep learning aims to extract hierarchical representations from large-scale data (e.g. images and videos) by using deep architecture models with multiple layers of non-linear transformations. With such learned feature representations, it becomes easier to achieve better performance than using raw pixel values or hand-crafted features.

Deep CNNs have recently demonstrated a great achievement in many computer vision tasks, such as object detection, object recognition, etc. Such deep CNN models have shown to appropriately characterize different variations within a large amount of data and to learn a discriminative non-linear feature representation. Furthermore, they can be easily generalized to other vision tasks by adopting and fine-tuning pretrained models through transfer learning [17, 19]. Thus, They provide a successful tool for different applications of FR by learning effective feature representations directly from the face images [17,18,19]. For example, DeepID, DeepID2, and DeepID2+ have been proposed in [28,21], respectively, to learn a set of discriminative high-level feature representations.

For instance, an ensemble of CNN models was trained in [21] using the holistic face image along with several overlapping/non-overlapping face patches to handle the pose and partial occlusion variations. Fusion of these models is typically carried out by feature

concatenation to construct over-complete and compact representations. Followed by [21], feature dimension of the last hidden layer was increased in [28], as well as, exploiting supervision to the convolutional layers in order to learn hierarchical and non-linear feature representations. These representations aim to enhance the inter-personal variations due to extraction of features from different identities separately, and simultaneously reduce the intra-personal variations. In contrast to DeepID series, an accurate face alignment was incorporated in Microsoft DeepFace [10] to derive a robust face representation through a nine-layer deep CNN. In [20], the high-level face similarity features were extracted jointly from a pair of faces instead of a single face through multiple deep CNNs for face verification applications. Since these approaches are not considered variations like blurriness and scale changes (distance of the person from surveillance cameras), they are not fully adapted for video-based FR applications.

Similarly, for the SSPP problems, a triplet-based loss function has been lately exploited in [22, 23, 24, 25, 19] to learn robust face embeddings, where this type of loss seeks to discriminate between the positive pair of matching facial ROIs from the negative non-matching facial ROI. A robust facial representation learned through triplet-loss optimization has been proposed in [23] using a compact and fast cross-correlation matching CNN (CCM-CNN). However, CNN models like the trunk-branch ensemble CNN (TBE-CNN) [22] and HaarNet [24] can further improve robustness to variations in facial appearance by the cost of increasing computational complexity. In such models, the trunk network extracts features from the global appearance of faces (holistic representation), while the branch networks embed asymmetrical and complex facial traits. For instance, HaarNet employs three branch networks based on Haar-like features, while facial landmarks are considered in TBECNN. However, these specialized CNNs represent complex solutions that are not perfectly suitable for real-time FR applications [29]. Moreover, autoencoder neural networks can be typically employed to extract deterministic non-linear feature mappings robust to face images contaminated by different noises, such as illumination, expression and poses [26,27]. An autoencoder network contains encoder and decoder modules, where the former module embed the input data to the hidden nodes, while the latter returns the hidden nodes to the original input data space with minimizing the reconstruction error(s) [26].

A generic auxiliary dataset containing faces of other persons can be exploited to perform domain adaptation [16], and sparse representation classification through dictionary learning [14]. However, techniques based on synthetic face generation and auxiliary data are more complex and computationally costly for real-time applications, because of the prior knowledge required to locate the facial components reliably, and the large differences between the quality of still and video ROIs, respectively.

## IV. DEEP LEARNING MODELS

### *A. Convolutional Neural Network*

The Convolutional Neural Networks (CNN) are very similar to ordinary Neural Networks. They are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. They still have a loss function (e.g. SVM/ Softmax) on the last (fully-connected) layer [30].

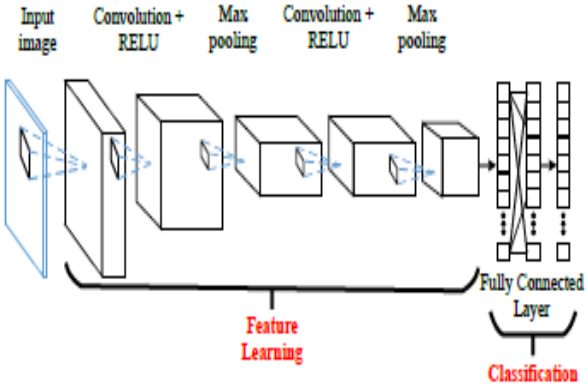


**Fig 1: Convolutional Neural Network (CNN)**

The CNN consists of multiple layers (see Fig. 1). Each layer takes a multi-dimensional array of numbers as input and produces another multidimensional array of numbers as output (which then becomes the input of the next layer). When classifying images, the input to the first layer is the input image (32×32), while the output of the final layer is a set of likelihoods of the different categories (i.e., 1 × 1× 10 numbers if there are 10 categories). A simple CNN is a sequence of layers, and every layer of a CNN transforms one volume of activations to another through a differentiable function. three main types of layers are used to build CNN architectures: Convolution (CONV) Layer, Pooling Layer, and Fully-Connected Layer and stacked these layers to form a full CNN architecture:

- INPUT [32×32] holds the raw pixel values of the image, in this case an image of width 32, height 32.
- CONV layer computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. This may

result in volume such as [32× 32 × 12] if we decided to use 12 filters.

- RELU layer applies an element wise activation function, such as the max(0; x) thresholding at zero. This leaves the size of the volume unchanged ([32× 32 × 12]).
- POOL layer performs a down-sampling operation along the spatial dimensions (width, height), resulting in volume such as [16 × 16 ×12].
- FC (Fully-Connected) layer computes the class scores, resulting in volume of size [1 ×1× 10], where each of the 10 numbers corresponds to a class score. As with ordinary Neural Networks and as the name implies, each neuron in this layer is connected to all the neurons in the previous volume.

Pooling layer (see Fig. 2) down samples the volume spatially, independently in each depth slice of the input volume. In this example, the input volume of size [224× 224 ×64] is pooled with filter size 2, stride 2 into output volume of size [112× 112 × 64] . Notice that the volume depth is preserved. The most common down sampling operation is max, giving rise to max pooling (see Fig. 2 down). That is, each max is taken over 4 numbers (little 2 × 2 square) [31].

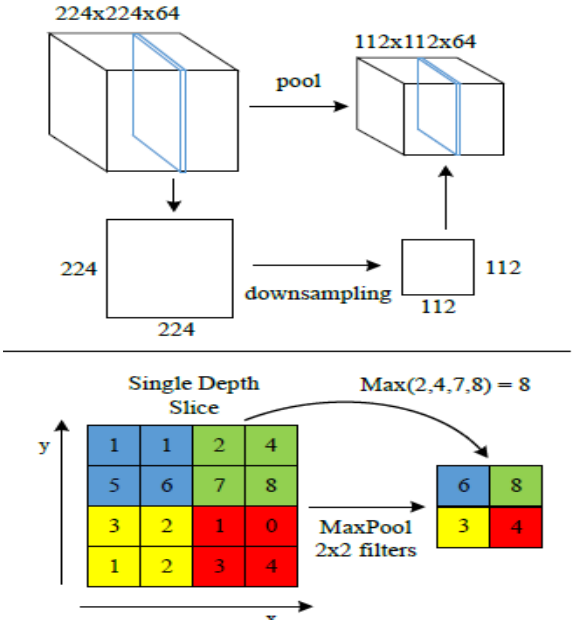


**Fig 2: Max pooling operation on feature map (2×2 window)**

The parameters in the CONV/FC layers have been trained with gradient descent so that the class scores that the CNN computes are consistent with the labels in the training set for each image [12].

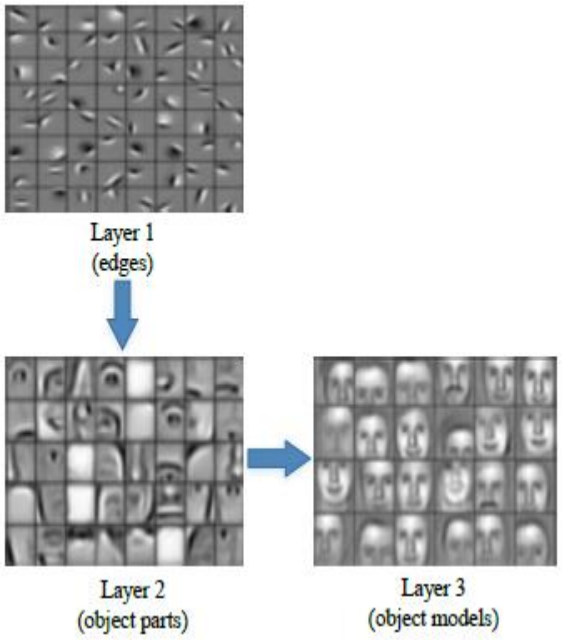


**Fig 3: Layers in Convolutional Neural Network (CNN)**

A CNN architecture is in the simplest case a list of Layers that transform the image volume into an output volume (e.g. holding the class scores) [32]:

- There are a few distinct types of Layers (e.g. CONV/FC/RELU/POOL are by far the most popular).
- Each Layer accepts an input 3D volume and transforms it to an output 3D volume through a differentiable function.
- Each Layer may (CONV/FC) or may not have (RELU/POOL) parameters.
- Each Layer may (CONV/FC/POOL) or may not have (RELU) additional hyper parameters.

In Fig. 3 we see representation of layers and filters in network for detecting face. First layer can detect basic edges. Second layer detects features from previous layers, thus it is able to detect more complex shapes like eye, nose or mouth. The third and last layer can detect whole faces.

### *B. Convolutional Restricted Boltzmann Machine:*

Huang et al. [33] propose to learn hierarchical features for face verification by using convolutional deep belief networks. The main contributions of this work are as follows:

i) a local convolutional restricted Boltzmann machine is developed to adapt to the global structure in an object class (e.g. face);

ii) deep learning is applied to local binary pattern representation [34] rather than raw pixel values to

capture more complex characteristics of hand-crafted features;

iii) learning the network architecture parameters is evaluated to be necessary for enhancing the multi-layer networks.

The convolutional restricted Boltzmann machine used in the proposed method is illustrated in Figure 4. It is reported that using the learned representations can achieve comparable performance with state of- the-art methods using hand-crafted features.

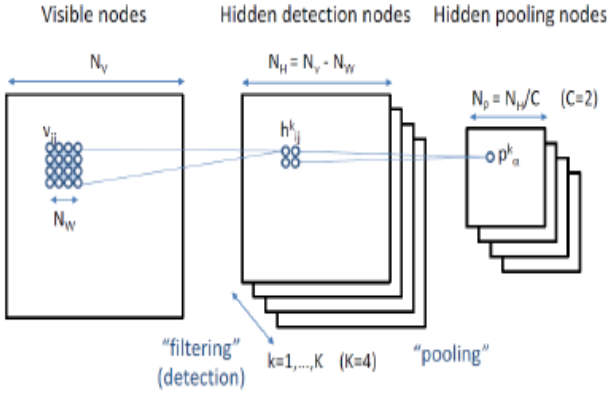


**Fig 4: Illustration of the convolutional restricted Boltzmann machine used in Huang et al.[33].**

### *C. 3D Face model*

Taigman et al. [35] propose a 3D face model based face alignment algorithm and a face representation learned from a nine-layer deep neural network.

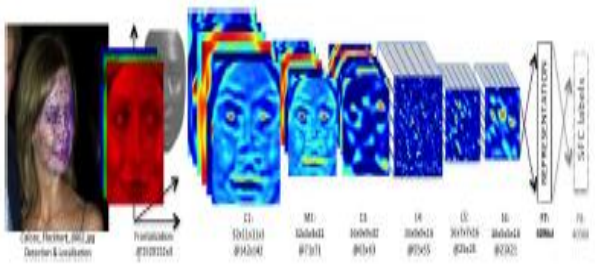

**Fig 5: Overview of nine layer deep neural network used in Taigman et al. [35]**

An overview of the architecture of the deep network is illustrated in Figure 5. The first three convolutional layers are used to extract low-level features (e.g. edges and textures). The next three layers are locally connected to learn a different set of filters for each location of a face image since different regions have different local statistics. The top two layers are fully connected to capture correlations between features captured in different parts of a face image. At last, the output of the last layer is fed to a K-way softmax which predicts class labels. The objective of training is to maximize the probability of correct class by minimizing the cross-entropy loss for each training sample.

It is shown that using the learned representations can achieve the near-human performance on the Labeled Faces in the Wild benchmark (LFW).

### *D. DeepID*

Sun et al. [36] propose to learn so-called Deep hidden IDentity features (DeepID) for face verification. In the feature extraction process. First, the local low level features of an input face patch are extracted and fed into a ConvNet [37]. Then, the feature dimension gradually decreases to 160 through several feed-forward layers, during which more global and high-level features are learned. Last, the identity class (among 10; 000 classes) of the face patch is predicted directly by using the 160-dimensional DeepID. Rather than training a binary classifier for each face class, Sun et al. simultaneously classify all ConvNets regarding 10000 face identities.

The advantages of this manipulation are as follows: i) effective features are extracted for face recognition by using the super learning capacity of neural networks;

ii) the hidden features among all identities are shared by adding a strong regularization to ConvNets.

It is reported that using the learned DeepID can achieve the near-human performance on the LFW dataset although only weakly aligned faces are used.

### *E. Joint Feature Learning:*

Lu et al. [38] develop a joint feature learning approach to automatically learn hierarchical representation from raw pixels for face recognition. the basic idea of the proposed method is First, each face image is divided into several non-overlapping regions and feature weighting matrices are jointly learned. Then, the learned features in each region are pooled and represented as local histogram feature descriptors. Lastly, these local features are combined and concatenated into a longer feature vector for face representation. Moreover, the joint learning model is stacked into a deep architecture exploiting hierarchical information.

## V. DEEP LEARNING ARCHITECTURES

Triplet-based loss optimization method allows to learn complex and non-linear facial representations that provide robustness across inter- and intra-class variations. CCM-CNN proposes a cost-effective solution that is specialized for still-to-video FR from a single reference still by simulating weighted CCM. TBE-CNN

and Haar-Net can extract robust representations of the holistic face image and facial components through an ensemble of CNNs containing one trunk and several branch networks.

In addition, to compensate the limited robustness of facial model in the case of single reference still, they were fine-tuned using synthetically-generated faces from still ROIs of non-target individuals. In contrast, CFR-CNN employed a supervised autoencoder CNN to generate canonical face representations from low-quality video ROIs. It can therefore reconstruct frontal faces that correspond to capture conditions of reference still ROIs and generate discriminant face representations.

## VI. CONCLUSION

The most recently proposed deep learning architectures for robust face recognition in video surveillance are focused to overcome the existing challenges in real-world surveillance unconstrained environments, the single training reference sample and domain adaptation problems with computational complexity is as a key issue to provide an efficient solution for real-time video-based FR systems has been studied.

As a result, the effectiveness of the proposed approaches of Face recognition techniques has been widely used in security systems and human-machine interaction systems. It is still a challenge for computer to automatically identify or verify a person due to large variations, e.g. illumination, pose and expression. Deep learning can utilize big data for training deep architecture models so as to obtain more powerful features for representing faces. In future, face recognition systems in smart cities will largely rely on hierarchical features learned from deep models.

## REFERENCES


[1] Zheng, J., Patel, V.M., Chellappa, R.: Recent developments in video-based face recognition. In: Handbook of Biometrics for Forensic Science, pp. 149–175. Springer (2017)

[2] Bashbaghi, S., Granger, E., Sabourin, R., Bilodeau, G.A.: Robust watch-list screening using dynamic ensembles of svms based on multiple face representations. Machine Vision and Applications 28(1), 219–241 (2017)

[3] Gomerra, M., Granger, E., Radtke, P.V., Sabourin, R., Gorodnichy, D.O.: Partially-supervised learning from facial trajectories for face recognition in video surveillance. Information Fusion 24(0), 31–53 (2015)

[4] Pagano, C., Granger, E., Sabourin, R., Marcialis, G., Roli, F.: Adaptive ensembles for face recognition in changing video surveillance environments. Information Sciences 286, 75–101 (2014)

[5] Bashbaghi, S., Granger, E., Sabourin, R., Bilodeau, G.A.: Dynamic ensembles of exemplarsvms for still-to-video face recognition. Pattern Recognition 69, 61 – 81 (2017)

[6] Barr, J.R., Bowyer, K.W., Flynn, P.J., Biswas, S.: Face recognition from video: A review. International Journal of Pattern Recognition and Artificial Intelligence 26(05) (2012)

[7] Matta, F., Dugelay, J.L.: Person recognition using facial video information: A state of the art. Journal of Visual Languages and Computing 20(3), 180 – 187 (2009)

[8] Dewan, M.A.A., Granger, E., Marcialis, G.L., Sabourin, R., Roli, F.: Adaptive appearance model tracking for still-to-video face recognition. Pattern Recognition 49, 129 – 151 (2016)

[9] Huang, Z., Shan, S., Wang, R., Zhang, H., Lao, S., Kuerban, A., Chen, X.: A benchmark and comparative study of video-based face recognition on cox face database. IP, IEEE Trans on 24(12), 5967–5981 (2015)

[10] Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: CVPR (2014)

[11] Bashbaghi, S., Granger, E., Sabourin, R., Bilodeau, G.A.: Watch-list screening using ensembles based on multiple face representations. In: ICPR, pp. 4489–4494 (2014)

[12] Kamgar-Parsi, B., Lawson, W., Kamgar-Parsi, B.: Toward development of a face recognition system for watchlist surveillance. PAMI, IEEE Trans on 33(10), 1925–1937 (2011)

[13] Kan, M., Shan, S., Su, Y., Xu, D., Chen, X.: Adaptive discriminant learning for face recognition.Pattern Recognition 46(9), 2497–2509 (2013)

[14] Yang, M., Van Gool, L., Zhang, L.: Sparse variation dictionary learning for face recognition with a single training sample per person. In: ICCV (2013)

[15] Mokhayeri, F., Granger, E., Bilodeau, G.A.: Synthetic face generation under various operational conditions in video surveillance. In: ICIP (2015)

[16] Ma, A., Li, J., Yuen, P., Li, P.: Cross-domain person re-identification using domain adaptation ranking svms. IP, IEEE Trans on 24(5), 1599–1613 (2015)

[17] Chellappa, R., Chen, J., Ranjan, R., Sankaranarayanan, S., Kumar, A., Patel, V.M., Castillo, C.D.: Towards the design of an end-to-end automated system for image and video-based recognition. CoRR abs/1601.07883 (2016)

[18] Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: CVPR (2012)

[19] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR (2015) [20]. Sun, Y., Wang, X., Tang, X.: Hybrid deep learning for face verification. In: ICCV (2013)

[20] Sun, Y.,Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: CVPR (2014)

[21] Ding, C., Tao, D.: Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans on PAMI PP(99), 1–14 (2017). DOI 10.1109/TPAMI.2017.2700390

[22] Parchami, M., Bashbaghi, S., Granger, E.: Cnns with cross-correlation matching for face recognition in video surveillance using a single training sample per person. In: AVSS (2017)

[23] Parchami, M., Bashbaghi, S., Granger, E.: Video-based face recognition using ensemble of haar-like deep convolutional neural networks. In: IJCNN (2017)

[24] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC (2015)

[25] Gao, S., Zhang, Y., Jia, K., Lu, J., Zhang, Y.: Single sample face recognition via learning deep supervised autoencoders. IEEE Transactions on Information Forensics and Security 10(10), 2108–2118 (2015)

[26] Parchami, M., Bashbaghi, S., Granger, E., Sayed, S.: Using deep autoencoders to learn robust domain-invariant representations for still-to-video face recognition. In: AVSS (2017)

[27] Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NIPS (2014)

[28] Canziani, A., Paszke, A., Culurciello, E.: An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678 (2016)
[29] BRITZ, D. Understanding convolutional neural networks. In: WILDML [Online]. 2015.Available at: http://www.wildml.com/2015/11/understanding-convolutional\-neural-networks-for-nlp/.
[30] TOBIAS, L., A. DUCOURNAU, F. ROUSSEAU,G. MERCIER and R. FABLET. Convolutional Neural Networks for object recognition on mobile devices: A case study. In: 23rd International Conference on Pattern Recognition (ICPR). Cancun: IEEE, 2016, pp. 3530–3535. ISBN 978-1-5090-4847-2. DOI: 10.1109/ICPR.2016.7900181.
[31] GUO, S., S. CHEN and Y. LI. Face recognition based on convolutional neural network and support vector machine. In: IEEE International Conference on Information and Automation (ICIA).Ningbo: IEEE, 2016, pp. 1787–1792. ISBN 978-1-5090-4102-2. DOI: 10.1109/ICInfA.2016.7832107.
[32] B. Huang, H. Lee, and E. G. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2518–2525.
[33] T. Ojala, M. Pietik¨ainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, vol. 29, no. 1, pp. 51–59, 1996.
[34] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[35] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10, 000 classes," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.
[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[37] J. Lu, V. E. Liong, G. Wang, and P. Moulin, "Joint feature learning for face recognition," IEEE Transactions on Information Forensics and Security, vol. 10, no. 7, pp. 1371–1383, 2015.