

Received May 16, 2020, accepted May 28, 2020, date of publication June 1, 2020, date of current version June 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999030

A Novel Neural Network Method for Face Recognition With a Single Sample per Person

MOHAMED ABDELMAKSOU¹, EMAD NABIL^{1,2}, IBRAHIM FARAG¹, AND HALA ABDEL HAMEED^{3,4}

¹Computer Science Department, Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt

²Computer Science Department, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia

³Information Systems Department, Faculty of Computers and Information, Fayoum University, Fayoum 63514, Egypt

⁴Khaybar Community College, Taibah University, Medina 42353, Saudi Arabia

Corresponding author: Emad Nabil (e.nabil@fci-cu.edu.eg)

ABSTRACT Face Recognition (FR) problem is one of the significant fields in computer vision. FR is used to identify the faces that appear over distributed cameras over the network. The problem of face recognition can be divided into two categories, the first is recognition with more than one sample per person, which can be called traditional face recognition problem. The second is the recognition of faces using only a Single Sample Per Person (SSPP). The efficiency of face recognition systems decreases because of limited references especially (SSPP) and faces taken in the Operational Domain (OD) different from faces in the Enrollment Domain (ED) in illumination, pose, low-resolution, and blurriness. This paper proposed a method that deals with all problems related to face recognition with SSPP. 3D face reconstruction is used to increase the reference gallery set with different poses and generate a design domain dictionary to overcome the problem of limited reference. Besides, the design domain dictionary is used to feed different deep learning models. Face illumination transfer techniques are utilized to overcome the illumination problem. Labeled Faces in the Wild (LFW) dataset is used to train Super-Resolution Generative Adversarial Network (SRGAN) to overcome the low-resolution problem. Deblur Generative Adversarial Network (DeblurGAN) is trained on the LFW dataset to overcome the problem of blurriness. The proposed method is evaluated using the Chokeypoint dataset and COX-S2V dataset. The final results confirm an overall enhancement in accuracy compared to techniques that use SSPP for face recognition (generic learning and face synthesizing approaches). Also, the proposed method outperforms of Traditional and Deep Learning (TDL) method accuracy, which uses SSPP for face recognition.

INDEX TERMS Face recognition, generative adversarial network (GAN), single sample per person (SSPP), illumination transferring, deep learning.

I. INTRODUCTION

Face Recognition (FR) is one of the significant fields in computer vision. FR systems are used to identify faces from videos recorded over several surveillance cameras distributed [1]. It is used in various applications including law enforcement, video monitoring, etc [2]. The main issue in Face Recognition is the limited number of references.

Many techniques deal with face recognition — the first one is a traditional technique that can identify faces by extracting

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Alatas¹.

features or landmarks from the image. An algorithm, for example, will investigate the relative scale, shape, and location of the eyes, nose, cheekbones, and use the features to scan for other photos with similar features such as Support Vector Machine (SVM) [3] and Principal Component Analysis (PCA) [4]. The second technique is dimensional recognition technique, which utilizes 3D sensors to capture a lot of information about face shape and use this information to recognize distinctive characteristics on the face, like a nose shape. This information enhances face recognition efficiency [5]. The third technique is skin texture analysis. It is another new pattern that utilizes visual details for the skin. The performance

of the face recognition technique increased when adding skin texture analysis [6]. The fourth technique is thermal cameras. It is a different shape of input data being taken for face recognition. The cameras will only recognize head shape by this process. In addition, it avoids the accessories of a person like makeup, hats, or glasses. The most important problem when we use thermal images for face recognition is that the datasets for face recognition are finite [7]. The problem of face recognition can be divided into two categories, the first is recognition with more than one sample per person, which can be called traditional face recognition problem. The second is the recognition of faces using only a Single Sample Per Person (SSPP). However modern face recognition techniques, like deep learning algorithms, achieved high accuracy, these techniques did not achieve good accuracy in the problems that have limited data. The reason is that deep learning techniques are data-hungry, which means to get high accuracy, you need more data.

Despite the current progress in machine learning and computer vision, designing a robust system for face recognition with SSPP persists a daunting issue in real-world monitoring applications. One fundamental problem is the visual field change between Enrollment Domain (ED), where faces are usually taken under controlled circumstances, and Operational Domain (OD) where faces are taken in uncontrolled circumstances with differences in illumination, pose, low-resolution, and blurriness, etc [8]. Another obstacle faced by the new face recognition techniques is the complexity of gathering samples. Few samples per person mean low effort to handle them, less cost to conserve and process them. Unfortunately, most techniques of face recognition depend significantly about the size and representativeness of the training set, and many of them would experience a severe drop in performance or fail to work if there is only one sample of training per individual. This is called a Single Sample Per Person (SSPP) problem. This problem achieves low performance and causes difficulties in the design of a robust FR system [9]. For most existing algorithms, this is very difficult due to the insufficient data of the training sample. Several strategies for dealing with SSPP problems have been suggested to enhance the efficiency of FR systems such as using multiple face representations [3], generic learning [10], creating artificial faces from the original reference [11], and deep learning [12].

A. PROBLEM STATEMENT AND MOTIVATION

In this paper, we have suggested a method that overcomes the problem of limited reference and overcomes the visual field change between images from the enrollment domain and operational domain with changes in lighting, blurriness, poses, etc. The first problem is the illumination problem; Face illumination transfer [13] technique is used to transfer the illumination of the image, which comes from the operational domain to illumination of the enrollment domain. The second problem is the blurriness image problem. Deblur Generative Adversarial Network (DeblurGAN) [14] is used to generate

a sharp image to overcome the problem of blurriness. The third problem is the pixel problem (low-resolution image), we use Super-Resolution Generative Adversarial Network (SRGAN) [15] to create a high-resolution photo to overcome the low-resolution photo problem. The limited reference is a big problem especially the SSPP problem, which is a big challenge in FR technology. So, we used 3D face reconstruction to reconstruct a three-dimensional (3D) face image from a two-dimensional (2D) face image to augment the reference gallery set with different poses.

The remainder of this paper is structured according to the following: Section II clarifies related works in these fields. Subsequently, the proposed method in section III. Section IV our implementation of the proposed method. Section V includes the experimental result. Eventually, conclusion and future works in section VI.

II. RELATED WORK

This section contains related-works for FR systems using a Single Sample Per Person (SSPP). Various techniques are proposed to increase the effectiveness of FR systems using a SSPP. They divided into multiple face representation technique, generic learning technique, generation of synthetic faces, and deep learning techniques. Below is a description of the technique.

A. MULTIPLE FACE REPRESENTATION

One effective technique for solving the SSPP problem in face recognition is to obtain distinguishing characteristics from images. Authors in [3], [16] implemented FR model which depends on different facial representations. They used various extraction techniques to patches in the ref image to generate many facial representations and to construct a collection of various examples-SVMs. This collection gives effectiveness against nuisance factors found in the examination models. In [17], Multiple Convolutional Neural Network process the facial picture to produce several poses different characteristics. Multiple face representation techniques can deal with for minor changes and are therefore not effective when dealing in realistic implementations which have a lot of variations (e.g., pose, blurriness and extreme illumination) [18].

B. GENERIC LEARNING

An earlier discovery to deal with visual field changes in face recognition models is using a generic set [19]. Several researchers have debated generic learning [20], [21], [22]. Authors in [19] Introduced an efficient generic learning approach for face recognition systems, using external information to determine the scatter matrix in-class for each person and apply this data to the reference images. Authors in [23] introduced a Robust Auxiliary Dictionary Learning (RADL) method which can extract characteristic information from a generic set through dictionary learning. Although the important improvements identified in using generic learning, many problems should be tackled. The generic intra-class variety might not be related to the gallery images, therefore it

may not be necessary to extract features from the generic set. Furthermore, A huge number of photos gathered from outside data that include redundant data that could lead to complex in implementing and reduce the ability to deal with intra-class variations [18].

C. GENERATION OF SYNTHETIC FACES

Increasing synthetically the reference gallery set is a different approach to balance the appearance changes in face recognition with SSPP. Shao *et al.* [24] Introduced FR method which depends on Sparse Representation-based Classification (SRC) that increases the dictionary by using a collection of artificial images created by measuring the difference between a pair of images. Authors in [25] increased the gallery of references by producing a collection of artificial images under camera-specific illumination circumstances to build a reliable FR system under examination circumstances. Blanz and Vetter [26] introduced a 3D Morphable Model (3DMM) for recreating a 3D image from a 2D image and subsequently synthesizing new face photos. Despite artificial faces can increase the performance of FR systems produced with an SSPP. These do not cover the spectrum of intra-class differences in the real application. A lot of synthetic faces should be generated for all possible capture circumstances in the operational domain. Without choosing representative photos from the external data and the generating reference, synthetic images could need complicated developments [18].

D. DEEP LEARNING FOR FACE RECOGNITION

The deep face recognition community follows the deep learning mainstream architectures, which were inspired by significant advances in object classification. DeepFace [27] was the first to use a nine-layer CNN in 2014, with multiple layers connected locally. It achieves an accuracy of (97.35%) on the Labeled Faces in the Wild (LFW) dataset [28]. In 2014, DeepId2 [29] uses contrastive loss and the accuracy on the LFW dataset reaches to (99.15%). FaceNet [30] used a large dataset in 2015 to train a GoogleNet [31]. It achieves an accuracy of (99.63%). In the same year, VGGFace [32] established a system to construct a large-scale Internet dataset. VGGFace reaches the accuracy of (98.95%). DeepFace, DeepId2, FaceNet, and VGGFace techniques achieve high accuracy because they solve traditional FR, which means more than one sample per person. Although the substantial progress in the deep face recognition architectures, training these architectures with SSPP is a big problem. Authors in [12] attempt to solve face recognition with (SSPP) using Deep CNN architecture. They suggest a combined Traditional and Deep Learning (TDL) system. They proposed expanding sample method to increase training samples and training CNN architectures. They achieve an accuracy of (74%) on the LFW dataset.

III. THE PROPOSED METHOD

The proposed method consists of two phases. The first phase is the design phase which discusses how to create synthetic

images using the 3D technique and generate a design domain dictionary. The second phase is the operational phase which discusses doing a set of preprocessing steps on face image that captured from video to overcome all problems related to face recognition then using deep learning approaches to recognize faces.

A. DESIGN PHASE

Only a single sample per person, So the need for increasing the reference for each person in a dictionary is essential. In this phase, we use the 3D face reconstruction to reconstruct a 3D face from 2D face photo to augment the reference gallery set with different poses. The used technique to reconstruct a 3D face is Position map Regression Network (PRN) technique [33]. This technique goes beyond all related works on several datasets with 3D face reconstruction, and it is straightforward with a very lightweight model. All those are accomplished through the intricate design of the 2D model of the 3D face shape and function of loss. Specifically, using a UV position map, that is a 2D photo that records the 3D coordinates of a full face while preserving the semantic sense at each UV location and training an encoder-decoder network. And therefore, from the 2D face image, reconstructing a 3D face. As shown in Fig.1, the reference set Contains one image per person and by using this image for each person can create a set of synthetic faces using Position map Regression Network (PRN) technique, And therefore the design domain dictionary contains for each person the reference and the Synthetic faces. The design domain dictionary is used in the operational phase to train different deep learning approaches.

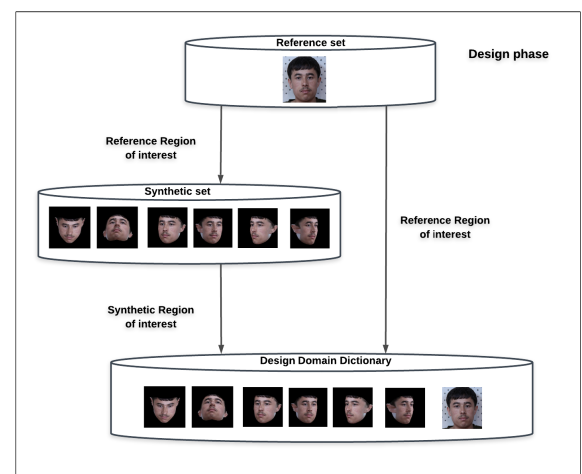


FIGURE 1. Design phase which generates a set of synthetic images and prepares a design domain dictionary that includes the reference and the Synthetic faces.

B. OPERATIONAL PHASE

We train different deep learning approaches to identify the person's face using a design dictionary that produced in the design phase. SRGAN is trained on the LFW dataset to

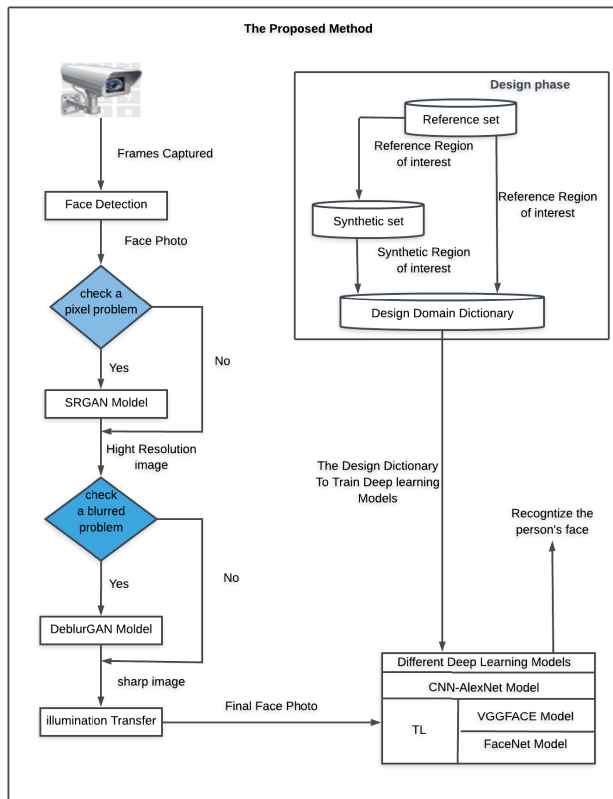


FIGURE 2. The proposed method block diagram using deep learning approaches.

overcome the problem of low-resolution faces. DeblurGAN is trained on the LFW dataset to overcome the problem of blurred faces. SRGAN and DeblurGAN work as a preprocessing step on the image that comes from the camera. The proposed method as illustrated in fig. 2 consists of the set of steps in the operational domain:

- 1) Detect faces from frames captured of video.
- 2) Check if the face photo less than size 96×96 (low-Resolution image). Then, the SRGAN model takes the face photo and generates a high-Resolution face photo.
- 3) Check if the face photo is a blurred image then the DeblurGAN model takes the face photo and generates a sharp face photo.
- 4) Transfer the illumination of face photo to illumination of enrollment domain.
- 5) Deep learning takes the final face photo to identify the person's face.

The details of each step are described below.

1) DETECT FACES FROM FRAMES CAPTURED OF VIDEO

It is the first stage in FR system is to separate the real face from the context of the image and differentiate between each face and another in the image. Even facial recognition algorithms must be capable to handle bad illumination conditions and different poses, such as angled or rotating faces. A lot of techniques are used for detecting faces such

as Dlib [34] library and Haar cascade classifier [35]. Either we should implement Dlib for detecting faces, using a hybrid of HOG (Histogram of Oriented Gradient) [36] & Support Vector Machine (SVM) [37] or OpenCV's Haar cascade classifier [35]. These techniques are trained on positive and negative images (which means certain pictures have faces and those that don't have). Dlib and OpenCV are capable of managing bad and inconsistent illumination and different facial poses like twisted faces. Dlib outperforms of Haar cascade classifier in precision. So, we will use Dlib in detecting faces.

2) SUPER RESOLUTION GENERATIVE ADVERSARIAL NETWORK (SRGAN)

GANs are deep neural network architectures and consist of two main networks (the Generator network and the Discriminator network). The architecture of the two networks consists mainly of the convolution layers, batch normalization and Parameterized Relu (PReLU). GANs focus on generating data from scratch. The aim of GANs is to generate new data that matches the training data distribution. It is like a game in which the Generator tries to generate some data from the distribution of probability and the Discriminator serves as a judge. Discriminator determines either the input comes from a true training dataset or fake generated data. The generator attempts to optimize data in order to match true training data. The discriminator guides the generator to generate realistic data. Discriminator and generator both learn concurrently and once the generator has been trained, it has enough knowledge of the distribution of training samples. Now the generator can synthesize new samples with very similar properties. SRGAN is used to produce higher resolution images and we will train it and use it to overcome the problem of low-resolution faces that come from the operational domain. The training procedure for SRGAN is shown in the following steps:

- 1) Processing the High-Resolution (HR) face photos for down-sampling Low-Resolution (LR) face photos. Now, we have HR and LR face photos for the training dataset.
- 2) Passing LR face images through Generator which up-samples and gives Super-Resolution face images.
- 3) Using a discriminator to distinguish the HR face images and back-propagate the GAN loss to train the discriminator and the generator.

SRGAN after training can take a low-resolution image and generate a high-resolution image. We use SRGAN after training in the operational domain. Check if the face photo less than size 96×96 (low-Resolution image). Then, the SRGAN takes the face photo and generates a high-Resolution Face photo.

3) DEBLUR GENERATIVE ADVERSARIAL NETWORK (DEBLUR-GAN)

The face image is checked if it is a blurred face or not. A single channel is taken of an image (presumably gray-scale)

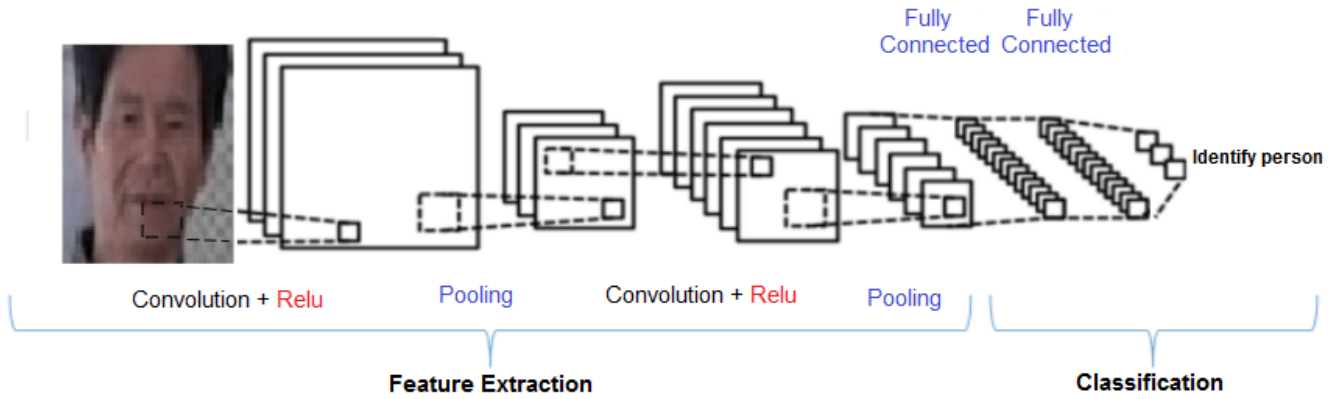


FIGURE 3. CNN architecture of AlexNet model.

and convolve it with 3×3 kernel defined as:

$$\text{TheLaplacianKernel} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

And then take the variance of the response. If the variance falls below a pre-defined threshold, then the image is considered blurry; otherwise, the image is not blurry.

DeblurGAN architecture consists of the network of generators that inputs the blurred image and generates a sharp image and the discriminating network to decide whether an input image is created artificially. We will train DeblurGAN to work on blurred face images. DeblurGAN after training can take a blur face image and generate a sharp face image. We use DeblurGAN after training in the operational domain. If the image that comes from the operational domain is blurred, then DeblurGAN takes the image to generate a sharp image to overcome the problem of the blurred face image.

4) TRANSFER THE ILLUMINATION OF FACE PHOTO TO ILLUMINATION OF ENROLLMENT DOMAIN

We use any image as a reference from the enrollment domain where usually images are taken within controlled circumstances, and all images from the Operational Domain (OD) apply the illumination transfer by using edge-preserving filters to overcome the problem of lighting. We use a novel image-based technique to transfer lighting from a reference image to a target image through the preserving filters [13]. According to this approach, just one reference image is required without any 3D geometric information of the target face. First, by using a Weighted Least Square (WLS) filter, decompose the light layers of the target and the reference pictures into large-scale. The final result is reached when the target picture's large-scale layer is exchanged with the reference picture's large-scale layer.

5) DEEP LEARNING APPROACHES

We studied the implementation of two forms of deep learning approaches for face recognition with SSPP:

CNN-AlexNet [38] and a transfer learning technique [39]. The design domain dictionary that is produced in the design phase is used to train these models.

The first deep learning approach is CNN-AlexNet: AlexNet [38] was chosen as an appropriate architecture for solving the problem of multi-classes classification. The overall structure is illustrated in Fig. 3, composed of facial photos as inputs. AlexNet consists of the layers of convolution and the max-layer and finished with two layers completely connected. The loss was measured using a softmax classifier, using multinomial logistic loss. The output of our network is to recognize the face. It consists of 2 fully connected layers with using the function Softmax described as:

$$f(o_j) = \frac{e^{o_j}}{\sum_{k=1}^K e^{o_k}} \tag{1}$$

where K is the number of classes and o_j is the j th output-vector value. The outputs of the Softmax are always in the range $[0,1]$ and add up to 1. The loss function is described such that making better decisions is equal to having a small loss while training. Included at the first completely connected layer is a Rectified Linear Unit (ReLU) layer. calculation of this operation is described as :

$$f(x) = \max(0, x) \tag{2}$$

where the function takes input with real value and threshold input to zero.

The second deep learning approach is Transfer Learning (TL) [40] in which a CNN is trained to learn characteristics for a wide domain then a modification of the classification function to know more technical features of a particular domain. The network parameters are moved from the wide domain to the particular domain. Moreover, Transfer Learning (TL) is also a technique that provides systems for applying the information gained from previous tasks to a new task related to the previous domain in some way. The transfer learning used is based on VGGFace [32], and FaceNet [30]. Table 1 shows the default picture sizes for the TL models.

TABLE 1. The input size of transfer learning models.

Models	Input Shape
VGGFace	224x224
FaceNet	160x160

IV. IMPLEMENTATION OF THE PROPOSED METHOD

In this section, we illustrate the dataset and explain the implementation of the two phases (design and operation phase).

A. DATASETS

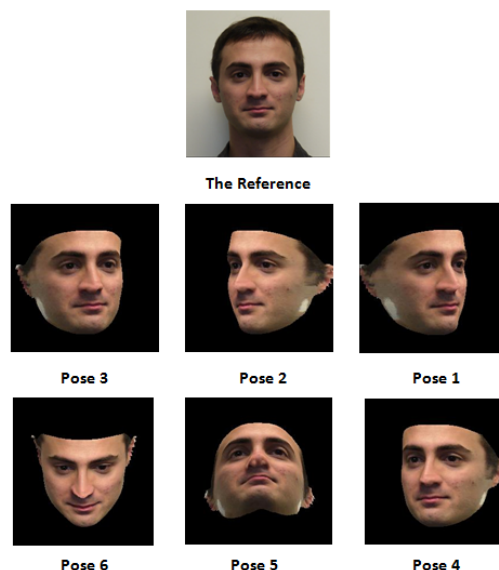
Extensive studies were carried out on two datasets which were publicly accessible – COX-S2V [41] and Chokepoint [42] to evaluate the FR system proposed under conditions of real-world surveillance. They include a high-resolution picture per subject which is taken under controlled circumstances and videos of low resolution for each subject taken under uncontrolled circumstances. Videos are taken through distributed cameras covering a variety of variations like pose, lighting, and scale.

The dataset of COX-S2V [41] involves thousand persons with one picture of high resolution and four low-quality video sequences for each video. Each person walks along an S-shaped path with light, pose and blur changes in each frame. The Chokepoint dataset [42] contains 25 people who walk through portal 1 and 29 who walk through portal 2. Portal one recording and Portal two recording two are a month apart. Just above the door is a surveillance setup with three cameras fixed and is used to monitor a person's entry during four sessions simultaneously. A total of 54 video clips and 64,204 face photos are included in the dataset. The appearance of the captured face varies in lighting, misalignment, and pose.

Labeled Faces in the Wild (LFW) [28] database is used to research the issues of face recognition problems and we use it to train Super-Resolution Generative Adversarial Network (SRGAN) and Deblur Generative Adversarial Network (DeblurGAN). It contains over 13,000 pictures of faces from the internet. every face was labeled with a person's name. It includes 5749 individuals with one or more images for each person.

B. IMPLEMENTATION OF DESIGN PHASE

We deal with a Single Sample Per Person (SSPP) problem. So for each person, we have only one image for each person in Chokepoint and COX-S2V databases. 3D Face reconstruction technique called Position map Regression Network (PRN) is used [33] to rebuild a 3D face from 2D face photo to augment the reference gallery set with different poses and make some transformations on the 3D face and choosing six images with different poses. So in the design domain dictionary for each person, we have 7 images (the reference and the synthetic faces) as shown in Fig.4. 3D Face reconstruction can overcome the problem of limited references.

**FIGURE 4.** Examples of synthetic faces with different poses using PRN technique.

C. IMPLEMENTATION OF OPERATION PHASE

First step: Different deep learning approaches are trained using the design domain dictionary that is produced in the design phase. The first approach is the proposed CNN approach AlexNet model [38]. We train AlexNet model by utilizing Adam optimizer and learning rate 0.0001. AlexNet model uses 100 epochs with a dropout rate of 0.30 to train. To get the best results on the datasets for the classification methods, we use regularization methods like dropout and normalization. The second approach is the transfer learning technique, we used two pre-trained models which are VGGFace [32], and FaceNet [30]. The pre-trained FaceNet model takes the image of a face and extract high-quality features and predict the vector representation of 128 elements of those characteristics, called the embedding face, but face embedding of VGGFace is a 2,048 length vector then we use the face embedding to recognize faces.

Second step: Super-Resolution Generative Adversarial Network (SRGAN) is trained on the LFW dataset to use it on low-resolution faces. First, we use the Pre-train VGG-19 model for training SRGAN. Second, we prepare the LFW dataset to train. For each image in the LFW training data with size 96×96 as a high-resolution image and downscaling size to 48×48 as a low-resolution image then fed the training dataset to SRGAN for training. After training, the model takes a low-resolution image and generates a high-resolution image. Thus, using SRGAN Model, we can overcome the problem of the low-Resolution image. As illustrated in fig. 5 in the left is the input of SRGAN (the low-resolution image). In the center is the output of SRGAN(the generated high-resolution image).

Third step: Deblur Generative Adversarial Network (DeblurGAN) is trained on the LFW dataset to use it on



FIGURE 5. SRGAN takes the low-resolution image (image in the left) and generates a high resolution (image in the middle).



FIGURE 6. DeblurGAN takes a blurred image (image in the left) and generates a sharp image (image in the middle).

blurred faces. First, we use the Pre-train VGG-19 model. Second, we prepare the LFW dataset to train. For each image in the LFW training data is a sharp image and blurring images using an Average Filter. Then fed the training dataset to DeblurGAN to train. After training, the model can take a blurred face image and generate a sharp image. So, using the DeblurGAN can overcome the problem of blurred faces as shown in fig. 6. In the left is the input of SRGAN (the blurred image). In the center is the output of DeblurGAN (the generated sharp image).

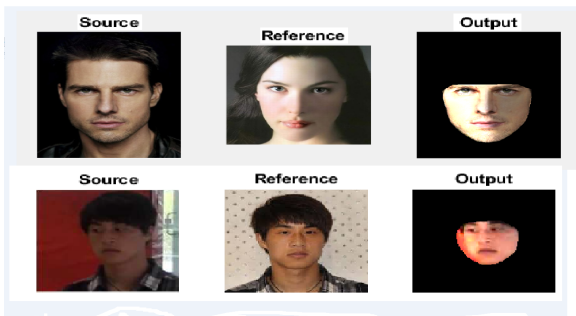


FIGURE 7. Edge-preserving filters technique converts the illumination of the operational domain to the illumination of the Enrollment domain.

Fourth step: Any image from the design domain is taking as a reference image and all images come from the operational domain as a target image. then Converting the illumination of the reference photo to the illumination of the target photo through Edge-preserving Filters. As shown in fig. 7. In the middle is the photo from the design domain as a reference photo. In the left is the photo from the operational domain as a target. And in the right is the result of converting the illumination of the photo that comes from the operational domain to the illumination of the design domain.

SRGAN, DeblurGAN, and Illumination transfer techniques are used as preprocessing steps on face image

before classification. Chokepoint and COX-S2V datasets are used to evaluate the proposed method. All videos that come from the operational domain in the two datasets are evaluated. The first step is to capture face from video. Then, check if the face photo less than size 96×96 (low-Resolution image) then the SRGAN takes the face photo and generates a high-Resolution Face photo. Then, Check if the face photo is a blurred image then the DeblurGAN takes the face photo and generates a sharp face photo. Then, illumination transfer techniques Transfer the illumination of face photo to illumination of enrollment domain and finally, Deep learning approaches take the final face photo to identify the person's face.

D. IMPLEMENTATION ENVIRONMENT

The experiments are carried out using API Keras v2.3.0 with TensorFlow v1.4 in the backend and we use Python (version 3.6) and the operating system used is Windows 7. The training and classification of the models are implemented on Intel Core i5-4570 @ 3.20GHz and GPU NVIDIA With 4GB, and 32 GB of RAM.

Moreover, SRGAN and DeblurGAN are trained using a server that uses the operating system Ubuntu 19.01 using Python, Keras, and Tensorflow. The server specification is intel Xeon Processor E5-2640 v2 @ 2.00 GHz \times 14, and 60 GB of RAM.

V. EXPERIMENTAL RESULTS

There are several face recognition issues with SSPP like pose, illumination, blurred and low-resolution image. The proposed method can deal with all these problems. The proposed method is evaluated using COX-S2V and Chokepoint dataset.

A. DESIGN PHASE RESULT

In the design phase for each face image in the Chokepoint and COX-S2V dataset, we generate six images with different poses. And therefore we have the design domain dictionary for COX-S2V and Chokepoint dataset. then the design domain dictionary is used to train different deep learning approaches.

B. OPERATIONAL PHASE RESULT

We train two network SRGAN and DeblurGAN to work as a pre-processing step of images. We train different deep learning approaches on the design domain dictionary that produced in the design phase on COX-S2V and Chokepoint datasets. When we used the proposed method with the proposed CNN approach AlexNet model, the result of Chokepoint and COX-S2V are 81.1%, 82.2%, respectively. But, when using the transfer learning technique, accuracy is increased. When using the proposed method with transfer learning on VGGFace, the result of Chokepoint and COX-S2V are 96.3%, 96.7%, respectively. When using the proposed method with transfer learning on FaceNet, the result of Chokepoint and COX-S2V are 98.7%, 98.5%, respectively. Furthermore, we made a comparison between our proposed

TABLE 2. Compare between the proposed FR method and FR approaches that deal with SSPP problem with Chokepoint and COX-S2V datasets.

Category	Method	Sub-Method	ChokePoint dataset	COX-S2V dataset
Generic Learning	ESRC [43]	SRC	80.2%	83.5%
Face Synthesizing	3DMM [26]	SRC	73.2%	75.7%
Generic Learning + Face Synthesizing	DSFS [18]	SRC	90.3%	92.5%
Deep Learning	TDL [12]	CNN	75.1%	74.8%
The proposed method	CNN-AlexNet		81.1%	82.2%
	TL	VGGFace [32]	96.3%	96.7%
		FaceNet [30]	98.7%	98.5%

TABLE 3. The average run time of the proposed method on Chokepoint and COXS2V datasets.

Technique	ChokePoint database		COX-S2V database	
	Number of Images	Run time(ms)	Number of Images	Run time(ms)
The proposed method	1000	285.71	2000	333.33

method with different techniques that deal with the SSPP problem on the same datasets as shown in Table 2. The generic learning technique [43] accuracy of Chokepoint and COX-S2V are 80.2%, 83.5%, respectively and Face Synthesizing technique [26] accuracy of Chokepoint and COX-S2V are 73.2%, 75.7%, respectively. But when using the generic learning and Face Synthesizing technique [18], the accuracy of Chokepoint and COX-S2V are 90.3%, 92.5%, respectively. Traditional and Deep Learning (TDL) technique [12] accuracy of Chokepoint and COX-S2V are 75.1%, 74.8%, respectively. As illustrated in Table 2, the proposed method that deals with face recognition systems with limited reference particularly a single sample per person achieves high accuracy up to 98.7% on the Chokepoint dataset and 98.5% on the COX-S2V dataset. We also compute the average run time of the proposed method using the pre-trained model FaceNet, which achieved the highest accuracy. As shown in Table 3, the average run time of the proposed method on the ChokePoint database is (285.71 ms) when testing 1000 image and on the COX-S2V database is (333.33 ms) when testing 2000 image.

On the other hand, our proposed method has some limitations which are:

- The training phase requires high computing resources and is time-consuming.
- In the operational phase, a lot of preprocessing steps on face image which require high computer resource.

VI. CONCLUSIONS AND FUTURE WORKS

This paper study the limited references problems in face recognition. We focus on Face Recognition with Single Sample Per Person (SSPP) problems. Faces captured under controlled circumstances in the enrollment domain different from taken under uncontrolled circumstances in the operational domain in illumination, pose, and blurriness. Therefore, the performance of FR systems decreases. To enhance face recognition system performance, we must increase the reference set by creating all possible artificial faces for all capture circumstances. Our proposed method deals with the

problem of limited references and overcomes the issue of a pose, illumination, blurriness, and low-resolution image.

We overcome the problem of a pose by using 3D Face Reconstruction to reconstruct a 3D face from 2D face image to augment the reference gallery set with different poses and overcome the problem of low-Resolution using Super-Resolution Generative Adversarial Network (SRGAN) And overcome the problem of blurriness using Deblur Generative Adversarial Network (DeblurGAN). Furthermore, overcoming the problem of illumination by making the illumination of the system constant by extracting the illumination from the enrollment domain and apply this illumination to the operational domain. We use different deep learning approaches. We run the proposed method with the CNN approach (AlexNet model) and transfer learning on the Chokepoint and the COXS2V dataset. The proposed method with transfer learning on FaceNet achieved high accuracy up to 98.7% on the Chokepoint dataset and 98.5% on the COXS2V dataset. we made a comparison between our proposed method with different techniques that deal with the SSPP problem, and the proposed method achieved high accuracy compared to techniques that use SSPP for face recognition (generic learning and face synthesizing approaches). Also, the proposed method outperforms of Traditional and Deep Learning (TDL) method accuracy, which uses SSPP for face recognition.

In the future work, we will deal with another important problem. The problem is when a person gets older, the facial features change. We will solve this problem using Generative Adversarial Network(GAN) by generating images of the same person in different years. Furthermore, we would like to enlarge the dataset by using artificial data augmentation techniques as used in [44].

REFERENCES

- [1] M. A. A. Dewan, E. Granger, G.-L. Marcialis, R. Sabourin, and F. Roli, "Adaptive appearance model tracking for still-to-video face recognition," *Pattern Recognit.*, vol. 49, pp. 129–151, Jan. 2016.
- [2] F. Matta and J.-L. Dugelay, "Person recognition using facial video information: A state of the art," *J. Vis. Lang. Comput.*, vol. 20, no. 3, pp. 180–187, Jun. 2009.

- [3] S. Bashbaghi, E. Granger, R. Sabourin, and G.-A. Bilodeau, "Dynamic ensembles of exemplar-SVMs for still-to-video face recognition," *Pattern Recognit.*, vol. 69, pp. 61–81, Sep. 2017.
- [4] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, "Discriminant analysis of principal components for face recognition," in *Face Recognition*. Cham, Switzerland: Springer, 1998, pp. 73–85.
- [5] J. Cook, V. Chandran, and C. Fookes, "3D face recognition using log-Gabor templates," in *Proc. Brit. Mach. Vis. Conf.*, M. J. Chantler, R. B. Fisher, and E. Trucco, Eds. Edinburgh, U.K.: British Machine Vision Association, Sep. 2006, pp. 769–778, doi: 10.5244/C.20.79.
- [6] J.-S. Pierrard and T. Vetter, "Skin detail analysis for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [7] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 72–114, Jul. 2003.
- [8] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 53–69, May 2015.
- [9] S. Bashbaghi, E. Granger, R. Sabourin, and G.-A. Bilodeau, "Ensembles of exemplar-SVMs for video face recognition from a single sample per person," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2015, pp. 1–6.
- [10] P. Zhu, M. Yang, L. Zhang, and I.-Y. Lee, "Local generic representation for face recognition with single sample per person," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 34–50.
- [11] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 296–301.
- [12] J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai, and F. Wang, "Deep convolutional neural network used in single sample per person face recognition," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–11, Aug. 2018.
- [13] X. Chen, M. Chen, X. Jin, and Q. Zhao, "Face illumination transfer through edge-preserving filters," in *Proc. CVPR*, Jun. 2011, pp. 281–287.
- [14] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.
- [15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [16] S. Bashbaghi, E. Granger, R. Sabourin, and G.-A. Bilodeau, "Robust watch-list screening using dynamic ensembles of SVMs based on multiple face representations," *Mach. Vis. Appl.*, vol. 28, nos. 1–2, pp. 219–241, Feb. 2017.
- [17] W. Abdalmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, R. Nevatia, and G. Medioni, "Face recognition using deep multi-pose representations," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.
- [18] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Domain-specific face synthesis for video face recognition from a single sample per person," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 757–772, Mar. 2019.
- [19] Y. Su, S. Shan, X. Chen, and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2699–2706.
- [20] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative multimaniifold analysis for face recognition from a single training sample per person," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 39–51, Jan. 2013.
- [21] Y. Gao, J. Ma, and A. L. Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, May 2017.
- [22] Z.-M. Li, Z.-H. Huang, and K. Shang, "A customized sparse representation model with mixed norm for undersampled face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2203–2214, Oct. 2016.
- [23] C.-P. Wei and Y.-C.-F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1722–1734, Jun. 2015.
- [24] C. Shao, X. Song, Z.-H. Feng, X.-J. Wu, and Y. Zheng, "Dynamic dictionary optimization for sparse-representation-based face classification using local difference images," *Inf. Sci.*, vol. 393, pp. 1–14, Jul. 2017.
- [25] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Synthetic face generation under various operational conditions in video surveillance," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4052–4056.
- [26] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
- [27] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.
- [28] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, E. Learned-Miller, A. Ferencz, and F. Jurie, Eds. Marseille, France, Oct. 2008.
- [29] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1988–1996.
- [30] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
- [31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [32] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. BMVC*, vol. 1, no. 3, 2015, p. 6.
- [33] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3D face reconstruction and dense alignment with position map regression network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 534–551.
- [34] J. Beel, B. Gipp, S. Langer, M. Genzmehr, E. Wilde, A. Nürnberg, and J. Pitman, "Introducing Mr. DLlib: A machine-readable digital library," in *Proc. 11th Annu. Int. ACM/IEEE Joint Conf. Digit. Libraries (JCDL)*, Jun. 2011, pp. 463–464.
- [35] P. I. Wilson and J. Fernandez, "Facial feature detection using Haar classifiers," *J. Comput. Sci. Colleges*, vol. 21, no. 4, pp. 127–133, Apr. 2006.
- [36] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 1491–1498.
- [37] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [40] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
- [41] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen, "A benchmark and comparative study of video-based face recognition on COX face database," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5967–5981, Dec. 2015.
- [42] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Proc. CVPR WORKSHOPS*, Jun. 2011, pp. 81–88.
- [43] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.
- [44] W. Al-Dhabyani, M. Goma, H. Khaled, and A. Fahmy, "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 1–11, 2019.



MOHAMED ABDELMAKSOU was born in Cairo, Egypt, in 1991. He received the B.S. degree in computer science from the University of Ain Shams, in 2013. He is currently pursuing the M.S. degree in computer science with the Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt.

He is also working in the smart village in Egypt. He is working in the machine learning field. He is interested in machine learning, image processing, convolution neural networks, the Internet of Things.



EMAD NABIL received the B.Sc. degree (Hons.) from the Computer Science Department, in 2004, the M.Sc. degree in soft computing and bio-inspired algorithms, in 2008, and the Ph.D. degree in optimization and machine learning, in 2012.

He is currently an Associate Professor with the Faculty of Computer and Information Systems, Islamic University of Medinah, Saudi Arabia. He is also with the Faculty of Computers and Artificial intelligence, Cairo University, Egypt. He is interested in machine learning, deep learning, optimization, bio-inspired algorithms, and bioinformatics.



IBRAHIM FARAG is currently a Professor of computer science, the Founder and the First Dean of the Faculty of Computers and Artificial Intelligence, Cairo University. He has published more than 300 articles in the fields of artificial intelligence and theoretical computer science. He is supervised more than 100 Master and Ph.D. thesis. His research interests are AI, machine learning, and theoretical computer science.



HALA ABDEL HAMEED received the bachelor's degree from the Faculty of Computers and Information Systems, Helwan University, in 2000, and the M.Sc. degree in 2006. Her master thesis title is E-Government in Egypt, theory and case study.

She is currently an Assistant Professor with the Faculty of Computers and Information, Fayoum University. She had a Ph.D. titled Replication in Distributed Real-Time Database Systems, in May 2012. She studied replication approaches in distributed real-time database systems and proposed a replication model DoMORE based on a hybrid employing of the two approaches of replication, implying an extension for both the theory and the implementation of them. Her current research interests include artificial intelligence, particularly, human cognitive functions, such as learning, reasoning, and machine learning techniques.

...