information

*Article*

# Face Identification Using Data Augmentation Based on the Combination of DCGANs and Basic Manipulations

**Sirine Ammar [1,2,*], Thierry Bouwmans [1] and Mahmoud Neji [2]**

[1] Laboratoire MIA, Université de La Rochelle, Avenue M. Crépeau, 17000 La Rochelle, France; tbouwman@univ-lr.fr
[2] Laboratoire MIRACL, Université de Sfax, Route de l'Aéroport, Sfax 3029, Tunisia; mahmoud.neji@fsegs.rnu.tn
* Correspondence: ammarsirine3@gmail.com

**Abstract:** Recently, Deep Neural Networks (DNNs) have become a central subject of discussion in computer vision for a broad range of applications, including image classification and face recognition. Compared to existing conventional machine learning methods, deep learning algorithms have shown prominent performance with high accuracy and speed. However, they always require a large amount of data to achieve adequate robustness. Furthermore, additional samples are time-consuming and expensive to collect. In this paper, we propose an approach that combines generative methods and basic manipulations for image data augmentations and the FaceNet model with Support Vector Machine (SVM) for face recognition. To do so, the images were first preprocessed by a Deep Convolutional Generative Adversarial Net (DCGAN) to generate samples having realistic properties inseparable from those of the original datasets. Second, basic manipulations were applied on the images produced by DCGAN in order to increase the amount of training data. Finally, FaceNet was employed as a face recognition model. FaceNet detects faces using MTCNN, 128-D face embedding is computed to quantify each face, and an SVM was used on top of the embeddings for classification. Experiments carried out on the LFW and VGG image databases and ChokePoint video database demonstrate that the combination of basic and generative methods for augmentation boosted face recognition performance, leading to better recognition results.

**Keywords:** generative methods; basic manipulations; data augmentation; FaceNet; SVM; face recognition; DCGAN



**Citation:** Ammar, S.; Bouwmans, T.; Neji, M. Face Identification Using Data Augmentation Based on the Combination of DCGANs and Basic Manipulations. *Information* **2022**, *13*, 370. https://doi.org/10.3390/info13080370

Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 5 July 2022
Accepted: 29 July 2022
Published: 3 August 2022



## 1. Introduction

Despite the exceptional efficiency of 2D and 3D recognition, face recognition based on DNNs has faced several challenges, such as the difficulty of collecting enough training images, because DNNs often require a large amount of data for effective learning. Generally, a large volume of data is useful to achieve high recognition accuracy. Because DNNs have powerful learning ability, they need various views of the face for each subject. However, obtaining such a dataset for a one class is not only impractical, it is very time-consuming. An insufficient amount of samples leads to over-parameterization and over-fitting issues, resulting in an obvious decline in the effectiveness of the learning outcomes. Moreover, it is often necessary to train samples of faces in different conditions of illumination, facial expressions, poses, and occlusion. To deal with the issue of scarcity or insufficiency of samples, an efficient solution is to use data augmentation techniques. The main purpose of data augmentation is to expand the diversity and size of the training database and expose the model to various aspects of data in order to guarantee that the model is evaluated on images that are not seen twice during training, achieving greater robustness, higher accuracy, and stable classification performance. However, there is a gap in the literature concerning research on the methodological and theoretical augmentation techniques for face recognition. Furthermore, there is a lack of studies on how to increase the size of small

datasets using augmentation. As a result, in the current work we focus on filling this gap. We propose a new data augmentation method and compare the effect of using various data augmentation techniques on face recognition accuracy. Specifically, we consider geometric transformations, image brightness changes, and application of different filter operations, as well as using DCGAN to generate new images from the original dataset to augment the amount of training samples.

A number of studies using deep learning methods have claimed high performance in a significant number of tasks. These include image classification [1], natural language processing [2], and text classification [3]. These models use the Softmax function in the classification layer. However, there have been studies [4,5] conducted that consider SVM as an alternative to the Softmax function for classification. These studies have asserted that the use of SVM in an artificial neural network (ANN) instead of the Softmax function may improve recognition accuracy. In our work, the augmented training data are sent to FaceNet to extract the embedding and then classified using an SVM.

The present work is an extension of our preliminary work published at the ISVC conference in 2020 [6]. The main contributions of this paper can be summarized as follows:

- We propose a novel data augmentation technique in which DCGAN and basic manipulations are combined. We use the Wasserstein loss to replace the standard DCGAN cross-entropy loss to solve the problem of DCGAN training instability. We show that our model improves face recognition performance by considering the LFW dataset [7], VGGFace2 dataset [8], and ChokePoint video database [9].
- We demonstrate the benefits of the proposed augmentation strategy for face recognition by comparing our approach with approaches using only basic manipulations and only a generative approach.
- We show that the use of SVM instead of the Softmax function with a FaceNet model may improve face recognition accuracy compared to the other tested techniques.

The rest of the paper is organized as follows: in Section 2, we review the literature in the area of data augmentation and face recognition; Section 3 presents the proposed approach in detail; Section 4 discusses the qualitative and quantitative results of our proposed method; and finally, Section 5 contains our conclusions and future research directions.

## 2. Related Works

### *2.1. Image Data Augmentation Techniques*

Image augmentations are generally categorized as traditional or generative augmentation methods. Traditional data augmentation methods involve geometric transformations, random cropping, kernel filters, color space augmentation, and noise injection [10]. Ben Fredj et al. [11] proposed a data augmentation technique with challenging facial appearance conditions and random information perturbation. This method is based on face representation using an adaptive fusion of center loss and softmax loss from data with occluded and extensive noisy faces. Noh et al. [12] suggested utilizing noise as a regularizing method in a DNN, particularly for dropout. This method defines stochastic hidden units as deterministic hidden ones with injected noise. The principal focus of this technique is to achieve lower bounds when increasing the number of samples in a stochastic gradient descent iteration. Noise injection has proven to be very powerful for learning robust features by adding random noise in each of the previously ranked input variables according to its relevance level, as demonstrated by Moreno-Barea et al. [13]. Xu et al. [14] applied flipping and cropping to generate more training data, even integrating the original image of the face and its mirror sample to boost face recognition performance. Zhong et al. [15] introduced Random Erasing, in which the pixels of the randomly selected region of the image are erased with random values. This technique can complement commonly employed data augmentation technologies, including random flipping and cropping. It can synthesize new samples with different levels of occlusion to mitigate the issue of overfitting and increase robustness under occlusion. Wu et al. [1] introduced a number of data augmentation techniques, such as vignetting, lens distortion, and color casting,

which can improve the sensitivity of CNNs to colors that are caused by scene illuminants. Mohammadzade and Hatzinakos [16] proposed projecting a face image with an arbitrary expression into the expression subspace to create new expressions in images for each person. This procedure yields significantly accurate estimation of within-subject variability. The "PatchShuffle" technique developed by Kang et al. [17] generates new images by randomly swapping pixel values in an $n \times n$ sliding window in each mini-batch. Lv et al. [18] proposed five data augmentation techniques for face recognition, covering four synthesis landmarks (illuminations, hairstyles, poses, glasses) and landmark perturbation. Li et al. [19] introduced a data augmentation technique, dubbed Moment Exchange (MoEx), that encourages the model to use the moment information of latent features. More precisely, the moments of the learned features of one training image are exchanged for those of another. This approach effectively improves classification accuracy and robustness across several datasets, model architectures, and prediction tasks. Despite the easy implementation of these methods, they remain the subject of research due to the increased training time and higher memory use they require.

Generative models are able to generate new training data, resulting in better performance by classification models. Generative approaches include methods such as Neural Style Transfer (NST) and Generative Adversarial Networks (GAN). NST [20] synthesizes a new image by manipulating the sequential representations across a CNN such that the style of one image can be transferred onto another while maintaining its original content. While this facilitates the transfer of textures, color temperatures, and lighting conditions, it can lead to bias in the data. Furthermore, NST algorithms [21] are slow. GANs create samples similar to images from the original dataset. Data augmentation based on GANs can enhance model performance even though the generated images do not appear hyperrealistic. However, at present the training of GANs faces many challenges, such as their instability and need for massive amounts of training data. Bowles et al. [22] used GANs as a way to unlock additional information from a dataset. Yi et al. [23] used a conditional generative adversarial network (cGAN) to expand the training dataset by generating emotional face images. In addition to GANs, there are many other generative models; however, GANs lead the way dramatically in terms of both the quality of their results and their computation speed. Variational auto-encoding [24] is one of the techniques utilized in augmentation that can improve the quality of samples produced by GAN. Variational auto-encoders are able to learn a low-dimensional representation of the data with regard to feature space augmentation.

In our earlier work [6], we proposed a data augmentation technique based on DCGAN to expand the training data by generating synthetic images. We demonstrate that DCGAN presents a strategy for using convolutional layers in the GAN framework to produce higher resolution images and improve the face recognition accuracy. In this work, we propose to extend our previous work [6] by combining generative models and basic image manipulation for data augmentation. We compare the use of only basic image manipulations methods, only generative methods, and the combination of these two techniques. As a generative approach, we used a DCGAN to generate synthetic image data. This is a generative approach that shows remarkable performance in generating different face attributes such as glasses, different colors of hair, and smiles. The main idea behind our proposed data augmentation method is to integrate strategies. We applied traditional manipulations (filtering, brightness changes, and geometric transformations) to images produced by the generative approach (DCGAN) in order to simulate the changes that can occur with a face. We then added these images to the original training dataset as augmented data. Sometimes, only a small amount of samples is required to train a CNN; however, with more images, the CNN has better performance.

### *2.2. Face Recognition Techniques*

#### 2.2.1. Conventional Methods

Face recognition is one of the most interesting biometric techniques. Technologies have been rapidly developed that allow for significant improvement in the precision of face recognition. There are many methods for recognizing faces in many applications, such as identity verification, face identification, security, surveillance, access control, and more. Ammar et al. [25] proposed a review of the different local and global approaches used for re-identification of people. They highlighted the fusion of local and global characteristics such as shape, color, and texture features combined with soft biometric characteristics including face shape, hair color, skin tone, eye shape, and eye color in re-identifying people through their faces. Moreover, Ammar et al. [26] provided an up-to-date review of face recognition approaches, covering earlier works as well as more recent advances. Anzar et al. [27] introduced a new algorithm based on Wavelet-SIFT descriptors for partial face recognition. Bi-orthogonal wavelet was applied to obtain the Discrete Wavelet Transform of the images. The scale-invariant feature transform (SIFT) algorithm was then used on high–high (HH) and low–low (LL) sub-bands of the images. The results showed a significant boost in recognition accuracy and a reduction in error rates. Ghorbel et al. [28] extracted features from face images using the VLC and Eigenfaces techniques, and used the chi-square distance for face matching. Haar cascade classifiers have demonstrated better performance than LBP classifiers in face detection, as presented by Johannes and Armin [29]. They demonstrated that the Eigenfaces technique is better than Fisher faces technique and LBP histogram technique for face recognition. In 2016, Khoi et al. [30] proposed to evaluate two variants of LBP, namely, Rotation-Invariant Local Binary Pattern (RILBP) and Pyramid of Local Binary Pattern (PLBP), for face retrieval. They split the facial image into small regions using the Grid LBP technique then constructed a histogram of spatially enhanced features based on the LBP feature vectors. Their system was robust against increases in the size of the dataset, with no sudden fall in mean average precision (MAP). A local appearance-based method called LBP network (LBPNet) has been proposed in [31] which effectively contributes to the extraction of hierarchical data representations. Laure et al. [32] used robust LBP to extract facial features to address the problem of large variations in lighting, expressions, and poses. KNN was applied for classification. Kumar et al. [33] proposed a dense local graph structure (D-LGS) descriptor using a bilinear interpolation to increase the pixel density when producing the graphic picture. The Histogram of Oriented Gradient (HOG) descriptor is widely used in face recognition applications. A Most Similar Region Selection algorithm (MSRS) was proposed by Karaaba et al. [34] to cope with misalignment by selecting similar regions of two face images. A multi-HOG model was first used to construct a distance descriptor, then mean of minimum distances (MMD) and multilayer perceptron-based distance (MLPD) functions were employed for face recognition. In Arigbabu et al. [35], face images were pre-processed using noise removal and a bicubic interpolation resampling technique. The shape of the face image was described locally using both Laplacian edge detection and the Pyramid HOG (PHOG) descriptor for human gender recognition. Based on the idea of the effectiveness of correlation filters under both controlled and uncontrolled environments, Napoléan and Alfalou [36] adopted a method based on the VLC correlator and Local Binary Patterns (LBP) technique to optimize the efficiency of face identification under illumination variations. In order to filter face images and extract the edges, they used a specific Gaussian function in the VanderLugt correlator and then applied an adapted LBP-VLC method. Face Recognition based on tensor methods has become very popular recently. Lu et al. [37] used Tensor Robust Principal Component Analysis (TRPCA), which aims to recover a low tubal rank tensor and a sparse tensor from the sum. They demonstrated that exact recovery can be achieved by solving a tractable convex program which does not have any free parameters. They used these tensor analysis tools to recover face images (of the same person) with random noise as well as for face image denoising. TRPCA, proposed by Cai et al. [38], aims to recover the low-rank and sparse components both efficiently and accurately in order to handle

high multi-dimensional data. The authors developed a t-Gamma tensor quasi-norm as a non-convex regularization to approximate the low-rank component. This configuration better captures the tensor rank while simplifying the process. In [39], the author compares TRPCA with other state-of-the-art low-rank factorization techniques for semi-supervised and supervised face classification tasks.

#### 2.2.2. Deep Learning Methods

Today, deep learning methods have superseded traditional face recognition methods. Research interests have mainly focused on face classification and recognition with DNNs. CNNs are a form of DNN that have achieved remarkable success in these areas. Unlike traditional face recognition algorithms [40], CNN is considered a data-driven method. Furthermore, CNNs combine both feature extraction and classification into one framework [11,41]. Song et al. [42] introduced a pairwise differential Siamese network to find correspondences between occluded facial blocks and corrupted feature elements for DCNN models, resulting in a face recognition model that is robust against occlusions. FaceNet, suggested by Schroff et al. [43], learns how to map from a face image to a Euclidean space embedding, in which distances between the embeddings directly correspond to a measure of face similarity. FaceNet directly trains 128-D compact embeddings using a loss function based on LMNN [44] as an online triplet mining method. The triplets incorporate two matching face thumbnails and a non-matching face thumbnail. The goal of the loss is to separate the positive pair from the negative by a distance margin. Hard-positive extraction techniques have been considered to support spherical clusters for the encodings of one person. FaceNet allows for much more prominent representational efficiency. Sphereface [45] presents an angular margin penalty to simultaneously impose extra intra-class compactness and inter-class separability. An additive Angular Margin Loss function has been proposed by Deng et al. [46], which can significantly increase the discriminating power of feature embeddings learned through CNNs. CNNs trained on 2D facial samples can be used successfully for 3D face recognition by refining the CNN with 3D facial scans, which allows for invariance to lightening/make-up/camouflage situations. In Tornincasa et al. [47], several linear quantities were employed as measures and relevant discriminant features were extracted from the query faces based on a differential geometry. Meanwhile, Dagnes et al. [48] presented an automatic method to calculate minimum optimized marker layouts to be used for capturing face motion. In 2015, DeepID3 networks [49] were proposed for face recognition; these were reconstructed from VGG stacked convolutions and the GoogLeNet inception layers. DeepID3 shows high performance on both face identification and verification. For face images captured under arbitrary illumination and pose conditions, Zhu et al. [50] proposed to use a deep network to recover the canonical frontal view. Each face was categorized as corresponding to a known identity by training a CNN. PCA was used with SVM on the network output for face verification. Taigman et al. [51] proposed to train a multi-class network on about four thousand identities in order to recognize faces. They employed a Siamese network to optimize the L1 distance between two face features. Their high accuracy obtained on LFW dataset [7] was the result of an ensemble of three networks using various color channels and alignments. The VGGNet-16 network presented by Simonyan et al. [52] achieved an accuracy of 98.95% using 2.6 million samples. Compared to DeepFace [51], which employs a single large CNN to learn features, Sun et al. [53] built their DeepID model for face representation to learn features from multiple CNNs using the network fusion technique. DeepID extracts features from images with various facial poses. Then, Sun et al. [54] developed DeepID2, which is an extension of DeepID. They employed a set of 25 nets, each taking a particular face patch as input. Then, 50 responses were combined to assess their performance under the LFW dataset [7]. A Joint Bayesian model [55] that usefully corresponds to a linear transformation in the embedding space has been proposed as well. To train the networks, the verification and classification losses are combined. The verification loss has the same concept as the triplet loss [44], as it aims to decrease the L2-distance between pairs

of faces belonging to the same identity and enlarge the distance between faces of various identities. The verification loss is similar to that applied in Wang et al. [56], in the sense that it compares only pairs of images in order to categorize images through semantic and visual similarity, while the triplet loss requires a relative distance constraint. In addition, deep learning-based approaches have addressed face recognition using GAN to extract more expressive features. Duan et al. [57] proposed a boosting GAN network in which unoccluded face images were generated from input occluded images for refined face recognition. The adversarial generator aims to obtain de-occlusion, coarse frontalization, and to preserve identity when both occlusions and pose variations exist simultaneously. However, this method is not generalized because it excludes certain types of occlusions, such as sunglasses, scarves, and masks. The main benefit of Deep Learning techniques is that they can be traine using massive amounts of samples in order to learn the high-level features for data representation, which can help to improve face recognition performance. However, the construction of such large datasets is beyond the capabilities of most academic groups.

In this paper, we chose FaceNet as a face recognition model. We applied FaceNet to images obtained by a combination of DCGAN and typical manipulations. As in other recent works that used deep networks [51,53], our method is a completely data-driven approach, which allows the representation to be learned directly from the face pixels.

## 3. Proposed Approach

Compared to our previous work [6], our proposed approach here is based on the combination of DCGAN and basic image manipulations for data augmentation to tackle the problem of scarce image data. Therefore, we decided to add synthetic images to the original face data. The face recognition system used here consists of several components, including fast and accurate face detection, face processing and cropping by computing facial landmarks using a Multi-task Cascaded Convolutional Neural Network (MTCNN), data augmentation by combining generative models and basic manipulations, face representation extraction, FaceNet training, and finally, applying SVM to classify and recognize faces in images and video streams, all of which is shown in the block diagram in Figure 1. Our proposed approach is carried out in the following steps: (1) MTCNN is applied for face detection; (2) the result of MTCNN is used as the input of the DCGAN to generate synthetic images; (3) basic manipulations are applied to the images created by the DCGAN; (4) images are added to the original data; and (5) FaceNet and SVM are applied for feature extraction and face recognition.

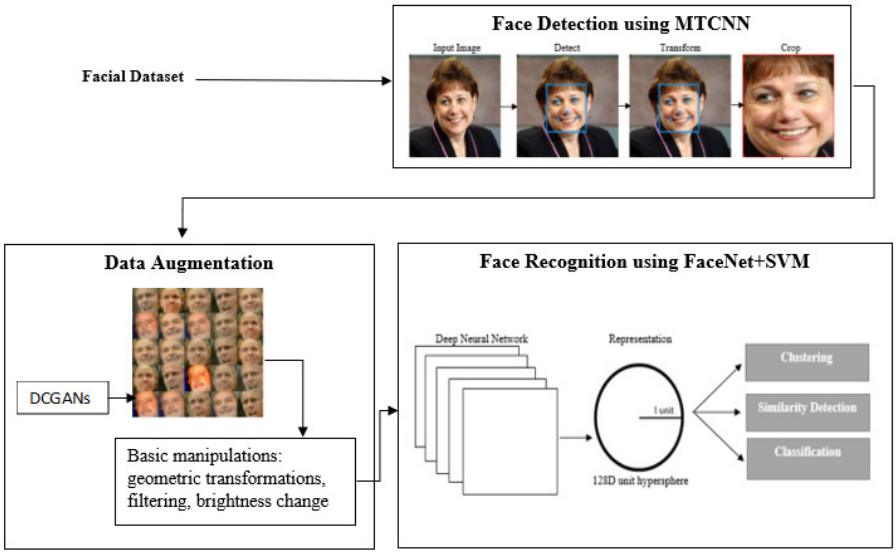



**Figure 1.** Overall architecture of the proposed approach.

### *3.1. Face Detection Using MTCNN*

As a first step, we performed the same preprocessing on all training and testing samples. We detected the location and extracted the canonical coordinates of the face from the input image or video frame using the MTCNN model [58]. MTCNN tries to perform a canonical face alignment by identifying the geometric structure of faces based on rotation, translation and scale. MTCNN is based on carrying out several tasks at the same time: (1) bounding box regression; (2) probability prediction that the face sample predicted by the MTCNN is a real face; and (3) facial landmark localization (location of eyes, mouth corners, and nose tip). MTCNN has several networks in a cascade: (1) P-Net processes the image at multiple resolutions and produces candidate face bounding boxes quickly; (2) R-Net refines the predictions and works as filter for selecting the high-accuracy candidate box; (3) O-Net further refines the predictions and generates the final bounding boxes. To remove overlapping bounding boxes, a Non-Maximum Suppression algorithm is applied. The network outputs the positions of the key features of the face. All input face images are cropped and further resized to 112 $\times$ 112 pixels according to the five facial points detected by MTCNN. A similar transformation is made depending on the position of the located keypoints, ensuring that all faces are cropped into images of a certain dimensions.

### *3.2. Data Augmentation Using DCGAN Combined with Basic Manipulations*

After applying face alignment and cropping, the extracted faces were passed through a DCGAN to synthesize new images. Basic manipulations are characterized by their simplicity and ease of implementation, however, they cannot reveal realistic face variations. While the generative approach can generate significant realistic face variations, it uses additional resources. To alleviate this issue, we trained the DCGAN [59] using the images from each class of the original dataset in order to learn the features of the faces and then artificially synthesize more facial images. We exploited the capacity of the DCGAN generator to synthesize more faces similar to the original faces in the training dataset, the idea being to simultaneously train two adversarial neural networks: the discriminator tries to discern whether the sample comes from the actual data distribution, while the generator aims to trick the discriminator by generating better samples. In the DCGAN training process, the discriminator aims to differentiate real samples from G(z), i.e., log(D(x)) + log(1 $-$ D(z)) should be maximized. Here, G(z) represents the generator, which takes as input a random noise z sampled from a uniform distribution, while D(.) represents the discriminator, which takes x as input, such as images, from the selected database or the output of G(z). At the same time, the generator aims to trick the discriminator by minimizing log(1 $-$ D(z)). A stable point in training is achieved when, after multiple steps, the discriminator is unable to differentiate between x and G(z). The generator takes a random noise vector as input sampled from a uniform distribution between ($-$1, 1), followed by a fully connected layer containing 8192 neurons and resized to the dimension of 4 $\times$ 4 $\times$ 1024, then four transposed convolutional layers with a stride of 2 and padding, which results in channel reduction and an upsampling of the features by factor of two. The final output images have a of size 64 $\times$ 64 $\times$ 3. The discriminator, D(x), is the inverse of the generator. The input image with dimensions 64 $\times$ 64 $\times$ 3 is transmitted through four consecutive convolutional layers with final output dimensions 4 $\times$ 4 $\times$ 512. A Softmax function is implemented to convert output real logits from the last fully connected layer into the final class probabilities. A learning rate of 0.0002 is used when training with the Adam optimizer, with a momentum term of $\beta_1$ = 0.5. The batch size is fixed at 128 and the weights are initialized from the normal distribution with a standard deviation of 0.02. Finally, a Nash equilibrium is achieved when the output of the discriminator is 0.5. That is to say, the discriminator D can no longer judge whether the input is from real data or from generated data. However, DCGANs have multiple problems, such as unstable training and unconvergent generator loss function. In this work, to solve the problem of mode collapse and vanishing gradient we propose using the Wasserstein distance [60] to evaluate the distance between generated samples and actual samples. The benefit of the Wasserstein distance is that it can measure the

distance between two non-coincident parts and further enhance the stability of training compared with the original DCGAN. The use of Wasserstein loss allowed the DCGAN to generate high-resolution synthetic faces with a high level of detail. Next, we applied basic manipulations, including geometric transformations (rotation and translation), brightness change, and filter operations (Gaussian filter, median filter, mean filter, and bilateral filter) to the DCGAN-generated images in order to enlarge the training data in a more efficient way. FaceNet was used to extract face embeddings, which were later classified using SVM model.

### *3.3. Face Recognition Using FaceNet Combined with SVM*

Our face recognition system integrates FaceNet with SVM for facial embedding feature extraction and classification, respectively. FaceNet [43] was used to obtain facial features and determine whether there was any matching between the input faces and a non-matching face with triplet-based loss function [54]. Figure 2 shows the structure of the FaceNet model, in which the face image is inserted into a DCNN which learns the features directly from the face image pixels, followed by an L2 normalization layer to finally obtain a 128-byte vector which results in the face image represented by the face embedding. Using an input face image, this model extracts 128-D vectors as face representations, which are then used to cluster faces in an efficient way. Unlike other face representation models, this embedding technique has the advantage that a larger distance between two face representations means that the faces are probably of different people. Training of the network requires a face triplet, a face image of the target person, a test face image of the target person, and a face image of a different person. This advantage facilitates the detection of similarities, grouping, and classification compared to other face recognition methods in which the Euclidean distance between features is not important. The network is trained in such a manner that the squared L2 distances between two embeddings correspond to face similarity. Faces of the "same" person have close distances and faces of different people have great distances. After this encoding has been generated, the distance between the two encodings is thresholded for face verification. To obtain the face embedding, FaceNet [43] generally uses two architectures based on CNNs. The first category adds $1 \times 1 \times$ d convolutional layers between the standard convolutional layers of the Zeiler and Fergus [61] architecture, then obtains a 22-layers NN1 model. The second category consists of an Inception model based on GoogLeNet [62]. Figure 3 represents the network structure of an Inception module. It contains four branches, from left to right. It employs convolution with $1 \times 1$ filters as well as $3 \times 3$ and $5 \times 5$ filters and a $3 \times 3$ max pooling layer. Each branch employs a $1 \times 1$ convolution to decrease time complexity. Finally, FaceNet employs the triplet loss function to train the model, which is used to minimize the distance between an anchor and a positive sample of the same person and to maximize the distance between the anchor and a negative sample.

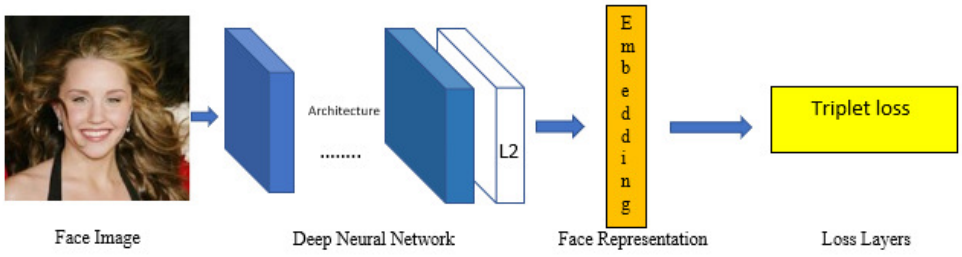



**Figure 2.** FaceNet model architecture: FaceNet consists of a batch input layer and a deep CNN (DCNN) followed by L2 normalization, which provides face embedding. Finally, the triplet loss is calculated during training.

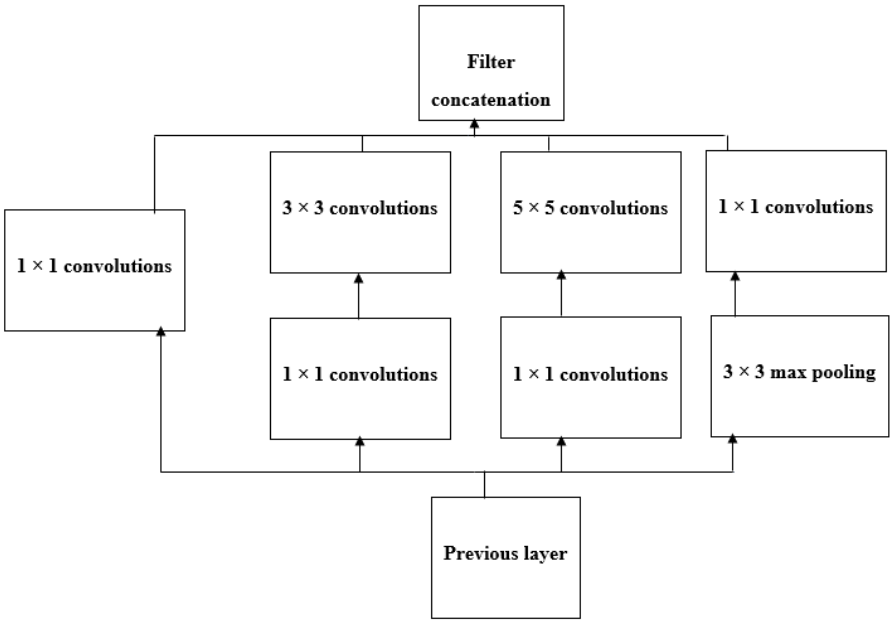



**Figure 3.** Inception module.

In this work, we propose to modify the original FaceNet networkby using SVM instead of the last fully connected layer, followed by L2 normalization. We propose to remove the last fully connected layer in a CNN and replace it with an SVM. As SVM is an efficient supervised learning algorithm for classification, this alteration creates a combined architecture with SVM for the facial recognition task. SVM is a very fast machine learning algorithm for multiclass classification, and it can be used for large data sets. Althought he FaceNet model can be used as part of the classifier itself, in our work we used the FaceNet model to pre-process a face in order to create a face embedding that could be employed as input in our classifier model. This latter approach was favoured because the FaceNet model is both large and slow to generate a face embedding.

## 4. Results and Discussions

In this paper, we propose to combine DCGANs and basic image manipulations as data augmentation techniques for our face recognition approach based on FaceNet + SVM. We evaluated the proposed method on two face image datasets, the LFW dataset [7] and VGGFace2 dataset [8], as well as on a video face dataset, the ChokePoint dataset [9]. We proposed to add 100, 250 and 500 generated images per one class. We compared our proposed approach with three popular face recognition algorithms: Principal Component Analysis (PCA), Tensor Robust Principal Component Analysis (TRPCA) [38], and Local Binary Pattern Histograms (LBPH). PCA is often employed for dimensionality reduction. It keeps only the values which contribute the most to variance in order to compress the dataset. It decomposes the covariance matrix to obtain the principal components (i.e., the eigenvectors) of the data and their corresponding eigenvalues. TRPCA [38] plays a critical role in handling high multi-dimensional datasets, aiming to recover the low-ranked and sparse components both accurately and efficiently. The LBPH method is based on the LBP, which is considered a texture description method. Using a face image, the histogram features are extracted from the occurrences of the LBP codes for texture categorization. Then, classification is carried out by finding the similarity between histograms. Additionally, we compared our approach with the work of Pei et al. [63], which was based on standard data augmentation techniques, including filtering, geometric transformations (rotation, translation, . . . ), and brightness changes along with a VGG-16 network for face classification. Furthermore, we compared our work with our own previous work [6] based on DCGANs for data augmentation and FaceNet + SVM for face recognition. In our experiments, we began by identifying faces in images, then moved on to identifying faces in videos.

### *4.1. Datasets*

#### 4.1.1. Labeled Faces in the Wild (LFW) Dataset

The LFW dataset is the standard dataset for face verification and recognition. This dataset contains 13,233 facial images of 5749 subjects. It includes several challenges, such as varying face poses, expressions, and illumination conditions as well as partial occlusion. In this dataset, only 1680 subjects out of a total of 5749 identities have more than one facial image. A subset of the database containing 3137 pictures obtained from 62 identities was employed in our experiments by choosing only those subjects with 20 or more images [7].

#### 4.1.2. VGGFace2 Dataset

The VGGFace2 dataset contains 9000 identities. The distribution of faces for different subjects is varied, from 87 to 843, with a mean of 362 images for each subject. We did not perform experiments on the entire dataset for time reasons. We selected a subset of the dataset by randomly choosing 20 identities to assess the performance of our method. The selected subset consisted of eight women and twelve men, with a total of 7746 samples [8].

#### 4.1.3. ChokePoint Dataset

The ChokePoint video dataset was designed for the identification of people acquired in real-world surveillance environments. The faces contained in the database have variations in terms of their pose, sharpness, and lighting, as well as misalignment due to the automatic detection/localization of faces. The Chokepoint video dataset consists of 25 identities (nineteen men and six women) in Portal 1 and 29 identities (twenty-three men and six women) in Portal 2. We used Portal 1 for our experiments [9].

### *4.2. Evaluation*

#### 4.2.1. Quality of Generated Images

In this paper, we trained a DCGAN model on two facial datasets, the LFW dataset [7] and VGGFace2 dataset [8]. Figures 4 and 5 show samples of images generated by the DCGAN. As expected, the images appear realistic, although with occasional artifacts. More realistic images can be generated by increasing the number of epochs. In this experiment, the DCGAN produces images which are difficult to assess subjectively against real images. These augmented samples are very similar to the original images. The quality of the generated images improved over the course of 40 epochs. DCGANs represent a very effective tool to modify attributes of real faces such as gender, age, skin, and hair color, and allow for surprisingly realistic results. DCGAN-generated samples cannot be distinguished by the CNN discriminator in terms of whether they are real or not. Thus, we assumed that DCGAN-generated samples would have similar CNN features, function as similar images, and help to create a limited dataset. Then, we applied basic manipulations on the synthetic images produced by the DCGAN. Figure 6 shows an example of the basic image manipulations applied to images generated by DCGAN. We added 100, 250, and 500 synthetic images per class to the LFW dataset [7] and VGGFace2 dataset [8]. Our proposed method demonstrates the ability to synthesize realistic faces with DCGANs and basic manipulations, which can be used to improve face recognition tasks.

**Figure 4.** Generated images using DCGAN on LFW dataset [7] after 50 epochs.

**Figure 5.** Generated images using DCGAN on VGGFace2 dataset [8] after 50 epochs.

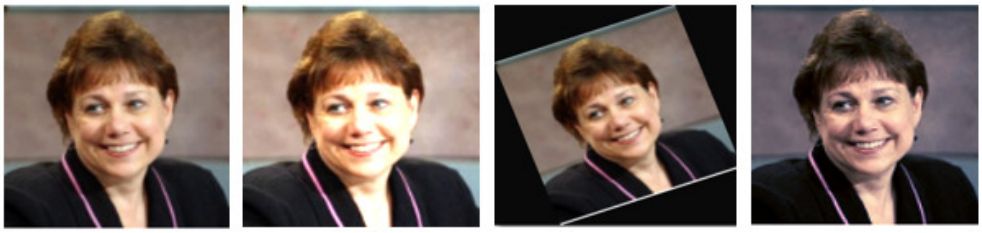


**Figure 6.** Example of basic manipulations applied on images from LFW dataset [7].

#### 4.2.2. Results

As shown in Table 1, the more training samples are added, the higher the accuracy of the model is. The results show that our face recognition approach achieves an accuracy of 64% and 61% with the LFW dataset [7] and VGGFace2 dataset [8], respectively. After a period of collecting more samples, the accuracy increases to 90% and 88%, respectively. Furthermore, when adding 500 samples per class, we achieve 94.5% accuracy on the LFW dataset [7] and 92.2% on the VGGFace2 dataset [8]. Table 2 shows that face recognition accuracy increases by 1.7% when adding only 100 images per class using ChokePoint dataset [9]. As expected, the face recognition accuracy increases as the amount of training samples increases.

**Table 1.** Face recognition accuracy with data augmentation using the proposed method.

| | **Number of Augmented Samples per Class** | | | |
|---|---|---|---|---|
| | +0 | +100 | +250 | +500 |
| LFW [7] | 0.64 | 0.79 | 0.90 | **0.945** |
| VGG [8] | 0.61 | 0.83 | 0.88 | **0.922** |

**Table 2.** Face recognition accuracy with data augmentation using the proposed method.

| | +0 | +100 |
|---|---|---|
| ChokePoint dataset [9] | 94.71% | **96.4%** |

Tables 3 and 4 show the results of the experiments using 62 classes from the LFW dataset [7] and 20 classes from the VGGFace2 dataset [8]. These experimental results demonstrate that our approach outperforms the traditional face recognition methods, namely, PCA, LBPH, and TRPCA. The results presented in Table 5, obtained with ChokePoint dataset [9], confirm the effectiveness of our approach. Our proposed method based on DCGAN and basic manipulations for data augmentation and FaceNet + SVM for face recognition has more advantages than the PCA, TRPCA and LBPH methods, which use a smaller number of samples.

**Table 3.** Recognition performance with different methods using 62 classes from LFW dataset [7].

| | |
|---|---|
| PCA method | 50% |
| Tensor RPCA [38] | 60.5% |
| LBPH method | 37% |
| CNN with filter operation augmentation method [63] | 65.4% |
| CNN with geometric transformations and brightness augmentation method [63] | 83.6% |
| FaceNet + SVM with filter operation augmentation method | 78.40% |
| FaceNet + SVM with geometric transformations and brightness augmentation method | 85.23% |
| FaceNet + SVM with DCGANs augmentation method [6] | 92.12% |
| Proposed approach (FaceNet + SVM + DCGANs + filter operation) | **93.4%** |
| Proposed approach (FaceNet + SVM + DCGANs + geometric transformations + brightness) | **94.5%** |

**Table 4.** Recognition performance with different methods using 20 classes from VGGFace2 dataset [8].

| | |
|---|---|
| PCA method | 40% |
| Tensor RPCA [38] | 41.5% |
| LBPH method | 32% |
| CNN with filter operation augmentation method [63] | 64.85% |
| CNN with geometric transformations and brightness augmentation method [63] | 74.21% |
| FaceNet + SVM with filter operation augmentation method | 79.6% |
| FaceNet + SVM with geometric transformations and brightness augmentation method | 78.94% |
| FaceNet + SVM with DCGANs augmentation method [6] | 91.83% |
| Proposed approach (FaceNet + SVM + DCGANs + filter operation) | **91.97%** |
| Proposed approach (FaceNet + SVM + DCGANs + geometric transformations + brightness) | **92.21%** |

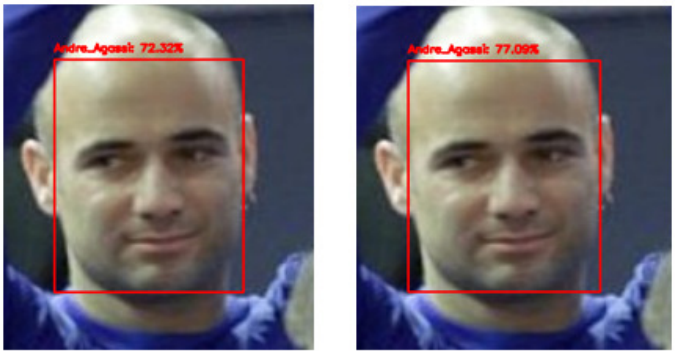
**Table 5.** Recognition performance with different methods using portal 1 from ChokePoint dataset [9].

| **Number of Augmented Samples per Class** | |
|---|---|
| PCA method | 50.4% |
| Tensor RPCA method [38] | 61.2 |
| LBPH method | 34.09% |
| CNN with filter operation augmentation method [63] | 69.83% |
| CNN with geometric transformations and brightness augmentation method [63] | 74.66% |
| FaceNet + SVM with filter operation augmentation method | 81.26% |
| FaceNet + SVM with geometric transformations and brightness augmentation method | 83.18% |
| FaceNet + SVM with DCGANs augmentation method [6] | 95.18% |
| Proposed approach (FaceNet + SVM + DCGANs + filter operation) | **95.8%** |
| Proposed approach (FaceNet + SVM + DCGANs + geometric transformations + brightness) | **96.4%** |

Moreover, we can observe that our proposed approach achieves better results than the work of Pei et al. [63], which was based on only basic data augmentation techniques (geometric transformations, brightness change, filtering and CNN for face recognition); see Tables 3 and 4. The number of augmented images did not change, at 500 images per face. The results shown in Table 5 confirm the effectiveness of our proposed approach on the ChokePoint dataset [9]. Again, the number of augmented images did not change, at 100 images per face.

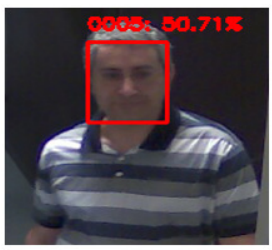
Additionally, the experimental evaluation demonstrates that a significant increase in accuracy can be obtained by combining DCGANs and basic manipulations for data augmentation and FaceNet + SVM for face recognition compared to only basic manipulations (geometric transformations, brightness change, filtering . . . ) as a data augmentation technique (see Tables 3–5). The obtained results show that our proposed approach using a DCGAN with a filter operation for data augmentation achieves higher accuracy than our previous work [6] based on only DCGANs, with a difference of 1.28% and 0.04%, respectively, for the LFW dataset [7] and VGGFace2 dataset [8] and 0.62% for the ChokePoint dataset [9]. Moreover, the results show improvement with our proposed approach using DCGAN and basic manipulations (geometric transformations, brightness change) as a data augmentation method, with a difference of 2.38% and 0.38% over our previous proposed method [6] for the LFW dataset [7] and VGGFace2 dataset [8], respectively, and 1.22% for the ChokePoint dataset [9]. These results show that augmentations can, in general, considerably improve the quality of face recognition systems, and that the combination of generative and basic manipulations performs better than the other tested techniques.

Figure 7 shows that the face *Andre_Agassi* from the LFW dataset [7] is recognized with 72.32% accuracy. However, the confidence is higher with data augmentation based on the combination of DCGAN and basic transformations, achieving 77.08%. In Figure 8a, we can see that the face prediction has only 50.71% confidence when using the ChokePoint dataset [9]; however, this confidence is higher when applying data augmentation with DCGAN and basic manipulations, achieving 91.63%, as shown in Figure 8b. The same is the case in Figure 8c,d with an increase of 1.8% when adding more samples per class. The results with the LFW database [7], VGG database [8], and ChokePoint database [9] show that the proposed approach can improve face recognition performance and lead to better recognition results.

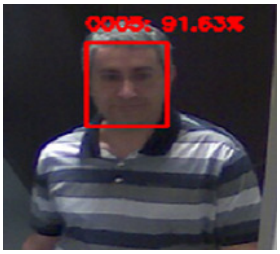



(**a**) Without data augmentation (**b**) With data augmentation

**Figure 7.** Face confidence using LFW dataset [7].




(**a**) Without data augmentation




(**b**) With data augmentation


(**c**) Without data augmentation


(**d**) With data augmentation

**Figure 8.** Face confidence using ChokePoint dataset [9].

## 5. Conclusions

Researchers use data augmentation to increase the size of the datasets used to train deep learning models. In this paper, we demonstrate that the combination of DCGAN and basic manipulations can generate data that approximate real face images, thereby both providing a larger data set for the training of large neural networks and improving the generalization ability of recognition models. Based on the augmented human face dataset, facial features were extracted using FaceNet and then classified using SVM. The effectiveness of the proposed method was demonstrated by various experiments and comparisons with frequently used data augmentation and face recognition methods. The proposed data augmentation method generates images realistic enough to boost the performance of face recognition systems.

Although the combination of approaches used here for augmentation demonstrates a good increase in accuracy, the basic approach is not far behind in terms of performance while requiring less time and hardware resources. Improving the quality of DCGAN-generated samples and evaluating their effectiveness on a broad range of datasets is a very important area for future work. DCGANs can be optimized by adjusting parameters such as batch size, momentum, and learning rate in order to generate more realistic and diverse face samples, which could further improve the accuracy of the results. Future work could include the use of Wasserstein loss with a gradient penalty to improve the quality of the generated images.

**Author Contributions:** Conceptualization, S.A.; writing—review and editing, T.B.; M.N. substantially revised the manuscript and supervised, and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wu, R.; Yan, S.; Shan, Y.; Dang, Q.; Sun, G. Deep image: scaling up image recognition. *arXiv* **2015**, arXiv:1501.02876.
2. Torfi, A.; Shirvani, R.; Keneshloo, Y.; Fox, E. Natural language processing advancements by deep learning: A survey. *arXiv* **2020**, arXiv:2003.01200.

3. Yang, Z.; Yang, D.; Dyer, C. Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.
4. Agarap, A.F. An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification. *arXiv* **2019**, arXiv:1712.03541v2.
5. Suguna, G.C.; Kavitha, H.S.; Sunita, S. Face Recognition System For Realtime Applications Using SVM Combined With FaceNet And MTCNN *Int. J. Electr. Eng. Technol. (IJEET)* **2021**, *12*, 328–335.
6. Ammar, S.; Bouwmans, T.; Zaghden, N.; Neji, M. Towards an Effective Approach for Face Recognition with DCGANs Data Augmentation. *Adv. Vis. Comput.* **2020**, 12509.
7. Huang, G.B.; Mattar, M.; Tamara, B.; Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In Proceedings of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Tuscany, Italy, 28 July–3 August 2008.
8. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A dataset for recognising face across pose and age. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, Xi'an, China, 15–19 May 2018.
9. Wong, Y.; Chen, S.; Mau, S.; Sanderson, C.; Lovell, B.C. Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition. In Proceedings of the IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops, Colorado Springs, CO, USA, 20–25 June 2011; pp. 81–88.
10. Kwasigroch, A.; Mikołajczyk, A.; Grochowski, M. Deep neural networks approach to skin lesions classification—A comparative analysis. In Proceedings of the International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland, 28–31 August 2017; pp. 1069–1074.
11. Ben Fredj, H.; Bouguezzi, S.; Souani, C. Face recognition in unconstrained environment with CNN. *Vis. Comput.* **2020**, *37*, 217–226. [CrossRef]
12. Noh, H.; You, T.; You, Mun, J.; Han, B. Regularizing deep neural networks by noise: Its interpretation and optimization. *Adv. Neural Inf. Process. Syst.* **2017**, 5109–5118.
13. Francisco, J.M.-B.; Fiammetta, S.; Jose, M.J.; Daniel, U.; Leonardo, F. Forward noise adjustment scheme for data augmentation. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 18–21 November 2018.
14. Xu, Y.; Li, X.; Yang, J.; Zhang, D. Integrate the original face image and its mirror image for face recognition. *Neurocomputing* **2014**, *131*, 191–199. [CrossRef]
15. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13001–13008. [CrossRef]
16. Mohammadzade, H.; Hatzinakos, D. Projection into expression subspaces for face recognition from single sample per person. *IEEE Trans. Affect. Comput.* **2013**, *4*, 69–82. [CrossRef]
17. Kang, G.; Dong, X.; Zheng, L.; Yang, Y. PatchShuffle regularization. *arXiv* **2017**, arXiv:1707.07103.
18. Lv, J.; Shao, X.; Huang, J.; Zhou, X.; Zhou, X. Data augmentation for face recognition. *Neurocomputing* **230**, *22*, 2017. [CrossRef]
19. Li, B.; Wu, F.; Lim, S.; Weinberger, K. On feature normalization and data augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 12383–12392.
20. Zheng, X.; Chalasani, T.; Ghosal, K.; Lutz, S. Stada: Style transfer as data augmentation. *arXiv* **2019**, arXiv:1909.01056.
21. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 27–30 June 2016; pp. 2414–2423.
22. Christopher, B.; Liang, C.; Ricardo, G.P.B.; Roger, G.; Hammers, A.; David, A.D.; Maria, V.H. GAN augmentation: augmenting training data using generative adversarial networks. *arXiv*, **2018**, arXiv:1810.10863.
23. Yi, W.; Sun, Y.; He, S. Data Augmentation Using Conditional GANs for Facial Emotion Recognition. In Proceedings of the Progress in Electromagnetics Research Symposium, Toyama, Japan, 1–4 August 2018.
24. Doersch, C. Tutorial on Variational Autoencoders. *arXiv* **2016**, arXiv:1606.05908.
25. Ammar, S.; Zaghden, N.; Neji, M. *A Framework for People Re-Identification in Multi-Camera Surveillance Systems*; International Association for Development of the Information Society: Lisbon, Portugal, 2017.
26. Ammar, S.; Bouwmans, T.; Zaghden, N.; Neji, M. From Moving Objects Detection to Classification And Recognition: A Review for Smart Cities. In *Handbook on Towards Smart World: Homes to Cities using Internet of Things Publisher*; CRC Press, Taylor and Francis Group: Boca Raton, FL, USA, 2017.
27. Anzar, S.M.; Amrutha, T. Efficient wavelet based scale invariant feature transform for partial face recognition. In *AIP Conference Proceedings*; AIP Publishing LLC: New York, NY, USA, 2020; Volume 2222, p. 030017.
28. Ghorbel, A.; Tajouri, I.; Aydi, W.; Masmoudi, N. A comparative study of GOM, uLBP, VLC and fractional Eigenfaces for face recognition. In Proceedings of the 2016 International Image Processing, Applications and Systems (IPAS), Virtual Event, Italy, 9–11 December 2016; pp. 1–5.
29. Johannes, R.; Armin, S. Face Recognition with Machine Learning in OpenCV Fusion of the results with the Localization Data of an Acoustic Camera for Speaker Identification. *arXiv*, **2017**, arXiv:1707.00835.
30. Khoi, P.; Thien, L.H.; Viet, V.H. Face Retrieval Based on Local Binary Pattern and Its Variants : A Comprehensive Study. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 249–258. [CrossRef]

31. Xi, M.; Chen, M.; Polajnar, D.; Tong, W. Local binary pattern network : A deep learning approach for face recognition. *IEEE ICIP* **2016**, *25*, 3224–3228.
32. Laure Kambi, I.; Guo, C. Enhancing face identification using local binary patterns and k-nearest neighbors. *J. Imaging* **2017**, *3*, 37.
33. Kumar, D.; Garaina, J.; Kisku, D.R.; Sing, J.K.; Gupta, P. Unconstrained and Constrained Face Recognition Using Dense Local Descriptor with Ensemble Framework. *Neurocomputing* **2020**, *408*, 273–284. [CrossRef]
34. Karraba, M.; Surinta, O.; Schomaker, L.; Wiering, M. Robust face recognition by computing distances from multiple histograms of oriented gradients. *IEEE Symp. Ser. Comput. Intell.*, **2015**, *7*, 10.
35. Arigbabu, O.; Ahmad, S.; Adnan, W.A.W.; Yussof, S.; Mahmood, S. Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor. *arXiv* **2017**, arXiv:1702.02537.
36. Napoléon, T.; Alfalou, A. Local binary patterns preprocessing for face identification/verification using the VanderLugt correlator. In *Optical Pattern Recognition*; SPIE: Bellingham, DC, USA, 2014; pp. 909–408.
37. Lu, C.; Feng, J.; Chen, Y.; Liu, W. Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5249–5257.
38. Shuting, C.; Luo, Q.; Yang, M.; Xiao, M. Tensor Robust Principal Component Analysis via Non-Convex Low Rank Approximation. *Appl. Sci.* **2019**, *9*, 7.
39. Liu, Y. *Tensors for Data Processing: Theory, Methods and Applications*, 1st ed.; Academic Press: Cambridge, MA, USA, 2021.
40. Qian, Y.; Gong, M.; Cheng, L. Stocs: An efficient self-tuning multiclass classification approach. In Proceedings of the Canadian Conference on Artificial Intelligence, Halifax, NS, Canada, 2–5 June 2015 ; pp. 291–306.
41. Wu, Z.; Peng, M.; Chen, T. Thermal face recognition using convolutional neural network. In Proceedings of the 2016 International Conference on Optoelectronics and Image Processing (ICOIP), Warsaw, Poland, 10–12 June 2016; pp. 6–9.
42. Song, L.; Gong, D.; Li, Z.; Liu, C.; Liu, W. Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network. In Proceedings of the 2019 International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
43. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
44. Weinberger, K.Q.; Blitzer, J.; Saul, L.K. Distance metric learning for large margin nearset neighbor classification. *J. Mach. Learn. Res. Adv. Neural Inf. Process. Syst.* **2009**, *10*, 207–244.
45. Liu, W.; Wren, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July; pp. 212–220.
46. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 4690–4699.
47. Tornincasa, S.; Vezzetti, E.; Moos, S.; Violante, M.G.; Marcolin, F.; Dagnes, N.; Ulrich, L.; Tregnaghi, G.F. 3D Facial Action Units and Expression Recognition using a Crisp Logic. *Comput. Aided Des. Appl.* **2019**, *16*, 256–268. [CrossRef]
48. Dagnes, N.; Marcolin, F.; Vazzetti, E.; Sarhan, F.R.; Dakpé, S.; Marin, F.; Nonis, F.; Mansour, K.B. Optimal marker set assessment for motion capture of 3D mimic facial movements. *J. Biomech.* **2019**, *93*, 86–93. [CrossRef]
49. Sun, Y.; Liang, D.; Wang, X.; Tang, X. Deepid3: Face recognition with very deep neural networks. *arXiv* **2015**, arXiv:1502.00873.
50. Zhu, Z.; Luo, P.; Wang, X.; Tang, X. Recover Canonical-View Faces in the Wild with Deep Neural Networks. *arXiv* **2014**, arXiv:1404.3543.
51. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 1701–1708.
52. Simonyan, K.; Zisserman, K. Very deep convolutional networks for large-scale image recognition. *arXiv*, **2014**, arXiv:1409.1556.
53. Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predecting 10,000 classes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitionhl, Columbus, OH, USA, 24–27 June 2014; Volume 23; pp. 1891–1898.
54. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep Learning Face representation by joint identification-verification. In Proceedings of the NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
55. Chen, D.; Cao, X.; Wang, L.; Wen, F.; Sun, J. Bayesian face revisited: A joint formulation. In Proceedings of the Computer Vision ECCV, Florence, Italy, 7–13 October 2012; pp. 566–579.
56. Wang, J.; Song, Y.; Leung, T.; Rosenberg, C.; Wang, J.; Philbin, J.; Chen, B.; Wu, Y. Learning grained image similarity with deep ranking. In Proceedings of the CVPR 2014: 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24–27 June 2014.
57. Duan, Q.; Zhang, L. Look more into occlusion: Realistic face frontalization and recognition with boostgan. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 214–228. [CrossRef]
58. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
59. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*, **2015**, arXiv:1511.06434.

60. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv* **2017**, arXiv:1701.078757.
61. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
62. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
63. Pei, Z.; Xu, H.; Zhang, Y.; Guo, M.; Yang, Y. Face recognition via deep learning using data augmentation based on orthogonal experiments. *Electronics* **2019**, *8*, 1088. [CrossRef]