

Article

Not peer-reviewed version

Synthetic Image Generation using Conditional GAN Provided Single Sample Face Image

[IQBAL MUHAMMAD ALI](#) , [Waqas Jadoon](#) , [Soo Kyun Kim](#) *

Posted Date: 8 April 2024

doi: 10.20944/preprints202404.0589.v1

Keywords: Convolutional Neural Network (CNN); Single Sample Per Person (SSPP); Conditional Generative Adversarial Network (CGAN)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Synthetic Image Generation Using Conditional GAN Provided Single Sample Face Image

Muhammad Ali Iqbal ¹, Waqas Jadoon ² and Soo Kyun Kim ^{1,*}

¹ Department of Computer Engineering, Jeju National University, Jeju 63243, Jeju Special Self-Governing Province, Republic of Korea

² Department of Computer Science, Comsats University Islamabad, Abbottabad Campus, Abbottabad, Pakistan; waqas_jadoon@cuiatd.edu.pk

* Correspondence: kimsk@jejunu.ac.kr

Abstract: Facial recognition systems are widely used systems around the globe. Their applications in real world scenarios like law enforcement, surveillance, ID card identification, information security, access control and many more has made them very important in today's world. These systems have shown promising results and excellent performance; however, these systems need tremendous number of facial images for training before they are implemented in the real-world scenarios. The problem arises when we face lack of training images due to which the performance of such systems decreases exponentially. The problem becomes even worse when there is only one image per subject for training our facial recognition system, the probe image might contain variation like illumination, expression and disguise which are hard to be accurately recognized. Generative models have achieved great success for handling the lack of training data by generating synthetic data that resembles the original data. We have created a new method by fine tuning the basic Conditional Generative Adversarial Network (CGAN) to generate synthetic facial images provided only single image per subject. These synthetic facial images contain six variations that are smile, anger, stare, illumination and disguise. We increased the size of the dataset and then applied a convolutional neural network for facial recognition, which improved the accuracy of our system from 0.76 to 0.99.

Keywords: Convolutional Neural Network (CNN); Single Sample Per Person (SSPP); Conditional Generative Adversarial Network (CGAN)

1. Introduction

Facial recognition has evolved as an emerging area in various fields like image processing, cognitive sciences, computer vision, machine learning and pattern recognition due to its important application in biometric verification technology. Human face can be referred from a video stream or a static image which can be recognized using human face databases that contains many different individuals. Humans has the natural ability to recognize faces effortlessly, however it is not easy for the machines to do such task so easily and distinctively due to the presence of complex features inside each human face, that makes every human distinct from one another [1]. For the past few decades machine learning and deep learning models have delivered some promising results in automatic face recognition. Automatic face recognition plays a vital role in critical tasks like identification of terrorists, military, general surveillance, identifying criminals, border security and immigration control.

Due to such important applications facial recognition has been always a hot research topic for the researchers. Automatic face recognition is a challenging task due to the involvement of several challenges such as illumination, pose, expressions, occlusion, and gender etc. Several methods have been developed for face recognition [2]. However, a good facial recognition system extracts key features from the face that are distinctive enough to separate the interclass variation as well as to reduce the intra class variation between the subjects at a low computational cost. Although a lot of progress have been made in recent years to develop a well-equipped facial recognition system, there is still a need to effectively address the solution regarding the challenges related to interclass variation (age of a person, certain appearances, race, gender), and intra class variations (illumination, disguise, expressions, occlusion). Deep learning has shown promising result in various fields in the past few

years. Several deep learning-based models have been used in facial recognition systems. However, these models are hard to train because they require a tremendous amount of data for training to perform accurately.

The main issue related to deep learning based facial recognition systems arises when there is a smaller number of training samples and sometimes there exists only a single training sample per person (SSPP). This makes it extremely challenging to recognize a query image with variant viewing condition like occlusion, illumination, pose and expression when having only one image per subject in the gallery set. The recognition systems fail to identify the facial variations due to limited amount of information in the gallery set. The SSPP problem is mainly handled using two basic approaches i.e., holistic, and local methods. Holistic methods use subspace models like Principal Component Analysis (PCA) [3], Linear Discriminant Analysis (LDA) [4] which require large training data which is the main problem encountered under SSPP scenario.

The local methods divide the face in to several local patches and ignore the holistic feature information among the training images. There are several methods that utilize a hybrid approach i.e., global and local facial feature extraction method but these methods are unpopular due to their high complexities and lower performance. Several synthetic image generation techniques have been developed to generate synthetic images by providing single sample face image, but these images were limited to only certain geometric changes, and they were biased toward limited poses. Moreover, the models trained on such synthetic images could not perform well on unseen test images.

Generative models come to a great rescue when dealing with the situation like lack of available training data. The most recent and the most powerful generative model is Generative Adversarial Network (GAN) [5]. The model comprises of two main CNN [6] based networks called generator and discriminator that are competing against each other in a min-max game in which generator and discriminator are to minimize the opponent reward and maximize its own reward. The generator produces a random noise 'x' which comprises of a random distribution i.e., noise. The discriminator contains the actual labelled data 'y'. The random noise is fed to the discriminator which compares it against the actual label 'y'. The classification error is calculated, and the weights of generator are updated. At each iteration the generator tries to produce more actual looking data while the discriminator makes sure to distinguish between the actual data and the fake data. At the end of the game the data 'x' produced by the generator is much closer to the actual data 'y'.

In this work we addressed the SSPP face recognition problem by generating virtual images using Conditional Generative Adversarial Network (CGAN) [7]. We trained six separate CGAN's which learned to generate six distinguished facial expressions that are smile, anger, stare, disguise, wearing, glasses, and illumination change. Unlike all the GANs our fine-tuned CGAN did not require any ground truth from the same facial image of the class, which means that our model is able to generate correct geometric facial representations of the single sample face image with accurate expressional variations on it without any additional knowledge. To test our model's generation accuracies, we build an additional Convolutional Neural Network (CNN) and trained it with our generated samples. After training, the CNN model was tested with the original images of the individuals which gave an accuracy of more than 99%, outperforming the current state of the art methods.

The rest of the paper is defined as section 2 explains all the related work related to SSPP and Image Generation Method. Section 3 Explains the details of the dataset. Section 4 explains the proposed model architecture and its working. Section 5 contains the results of experimentation and Section 6 concludes the paper.

2. Related Work

In literature the facial recognition process is broadly divided in to two categories: one is holistic feature extraction approach and the second one is the local feature-based approaches, the lateral approach outperforms the former by an accuracy of 60% [8]. Face recognition using Deep Convolutional Neural Network (DCNN) is currently the hot topic because of its excellent facial recognition accuracy

over traditional methods. Recent literature has reported the state-of-the-art work on facial recognition using CNN. For instance, In the paper [9] have used CNN to address the facial recognition under multiple samples and single sample per person. For multiple samples per person, they learned the activation vector from the fully connected layer using Visual Geometric Group (VGG) face feature extractor and normalized it with L2 normalization and in the end LDA is applied. In the paper [10] proposed a method which uses the local binary patterns as CNN's input and then use softmax regression function for face recognition. In the paper [11] the author presented joint collaborative representation with adaptive convolutional feature for Single Sample Problem (SSP). The method extracts the local regions of the query image using CNN. These regions have similar coefficient which helps to preserve all the local discriminative features, robust to different facial variation. In the paper [12] proposed a supervised auto encoder technique in which they mapped multiple versions of a face image to a canonical face and extracted features that preserved a similarity criterion. However, they only used few training images of size 32×32 which do not effectively tackle the facial variations of large set of query samples.

In the paper [13] the authors proposed sparse illumination transfer method which create illumination dictionary. The illumination transfer method only focused on dealing with the illumination changes on frontal face image but didn't focus on the actual shape of the face by extracting its global and local features. In [14,15] proposed patch-based method which extracted patch-based distribution of the training image and used a voting strategy to calculate the distances between path to patch manifolds. In [16] authors applied divide and conquer strategy. They divide the face image into a set of non-overlapping blocks. Each block is further divided into overlapped patches, assumed to lie in linear subspace. In [17,18] proposed single sample per person domain adaptive network technique. The method assumed that the images obtained in the gallery set are under stable shooting conditions. To deal with lack of training sample they used 3D face model and generated synthetic images to address different pose variations.

In the paper [19] the author construct feature dictionaries using both training and test images for under-sample face recognition in IoT based applications. In [20] proposed an iterative dynamic generic learning method which incorporates semi-supervised low rank representation framework for prototype recovery to learn variation dictionary for SSPP problem. In the paper [21] the author combined grayscale monogenic features and kernel sparse representation on multiple riemannian manifolds. The approach extracts regional face discriminability and co-occurrence distributions, fusing multiple kernels using kernel alignment and normalization. In the paper [22] the author proposes a binary coding method that is based on self organizing map, the method combines the self organizing map and the bag of features model to extract semantic features in the middle level from the facial images, they also utilize sift descriptor to obtain local features and then mapping them into semantic space using self organizing map.

In the paper [23] the author proposed an approach for 2D pose-variant face recognition using frontal view, they calculate a pose view angle and match the test images with rotated canonical face images, the test face is then warped to create a frontal view using landmark features, the distortion is also addressed by applying prior facial symmetry assumptions before matching with the frontal face. In the paper [24] the author proposed an integrated approach by combining feature pyramid and triplet loss technique, their approach reduce the computations by sharing the same backbone network. In the paper [25] the author proposed a probabilistic interpretable comparison approach which is a scoring approach proposed for biometric systems, the approach combines multiple samples using bayes theorem and offers probabilistic interpretation of decision correctness.

These approaches mainly generated virtual samples to address the SSPP, however they do not effectively address the adversarial problem related to the SSP that are different facial variations like anger, smile, occlusion or any other facial expression which exists in the query image and SSP is still an open research challenge.

Generative Adversarial Network (GAN) are popularly used for image synthesis, due to their capability in producing nearly realistic images which cannot be easily distinguished while viewing the image at first glance [26]. GANs are utilized for various purposes like texture synthesis, which enhances the resolution of images known as Super Resolution Generative Adversarial Network (SRGAN) [27]. In [28] the author up sampled low resolution images up to 4 times its original resolution. Adversarial loss and feature loss were considered as the main objective functions with feature maps obtained from pre-trained VGG 19 model. The loss function was calculated by computing the overall difference between the generated up sampled images and the ground truths.

In [29] developed Age Conditional Generative Adversarial Network (AGE-CGAN), the network consists of conditional generator and an encoder. The additional information containing the age information is kept in condition 'c' and a latent vector 'z' is constructed for the face image. Aging factor is calculated for the input image in the resultant generated image. In [30] the author developed an approach based on GAN called Disentangled Representation learning-GAN (DR-GAN). The encoder-decoder network architecture learns the representation of disentangled face images. These images are the input for the encoder and are output for the decoder. The encoder identifies features from the face image and the decoder generate synthesized images given with identity feature target pose and noise vector. The discriminator classifies on identity and poses to distinguish inter-class variations.

3. Proposed Model

The proposed CGAN [31] model comprises of two networks; the generator and the discriminator as described earlier. However, CGAN is slightly different from normal GAN as it takes input in pairs. The CGAN consist of a generator and a discriminator. The generator takes an input image and applies different convolutional filters on that image and yields an output image. The discriminator takes images in pair with input image paired with original target image and a paired image with the output image from the generator. It compares both paired images and distinguishes which image was produced by the generator, the weights of the generator are updated, and the generator produces a new image which is much better than the previous image produced. The process continues until the generator produces a synthetic image that is most realistic, and the discriminator is not able to distinguish between the real mage and the synthetic image.

3.1. Proposed Method

GANs are generative models that learn mapping from random noise vector 'c' to output image 'b', $G : c \rightarrow b$. In contrast, conditional GANs learn a mapping from observed image 'a' and random noise vector 'c', to 'b', $G : \{a, c\} \rightarrow b$. The generator 'G' is trained to produce output that cannot be distinguished from real images by an adversarial trained discriminator, 'D', which is trained to do as well as possible at detecting the generator's "fakes".

3.2. Objective Function

In a conditional Generative Adversarial Network (GAN), the goal is to train a generator (G) to produce realistic images conditioned on some input. This input could be an image (a) and/or a random noise vector (c). The discriminator (D) is trained to distinguish between real and generated images.

The objective function of a conditional GAN can be described as follows:

$$L_{CGAN}(G, D) = \mathbb{E}_{a,b}[\log D(a, b)] + \mathbb{E}_{a,c}[\log (1 - D(a, G(a, c)))], \quad (1)$$

where $\mathbb{E}_{a,b}$ denotes the expectation over pairs of real images (a) and their corresponding labels (b), and $\mathbb{E}_{a,c}$ denotes the expectation over pairs of real images (a) and generated images ($G(a, c)$) conditioned on random noise vector (c).

The generator (G) aims to minimize this objective, while the discriminator (D) aims to maximize it. Hence, the optimal generator (G^*) is obtained by solving the minimax problem:

$$G^* = \arg \min_G \arg \max_D L_{CGAN}(G, D). \quad (2)$$

In practice, the $L1$ distance measure is often preferred over $L2$ because it results in less blurring effect in the generated images. Therefore, the final objective of training a conditional GAN includes the $L1$ distance term, given by:

$$G^* = \arg \min_G \arg \max_D L_{CGAN}(G, D) + \lambda L_{L1}(G), \quad (3)$$

where λ is a hyper-parameter controlling the importance of the $L1$ loss term.

3.3. Proposed Generator and Discriminator Network Architecture

The proposed Conditional Generative Adversarial Network (CGAN) model consists of a generator and a discriminator, as depicted in Figure 1 and Figure 2. The generator network comprises an encoder and a decoder. The encoder accepts an image as input and applies a series of convolutional layers. This process results in down-sampling of the image, as illustrated in Figure 2.

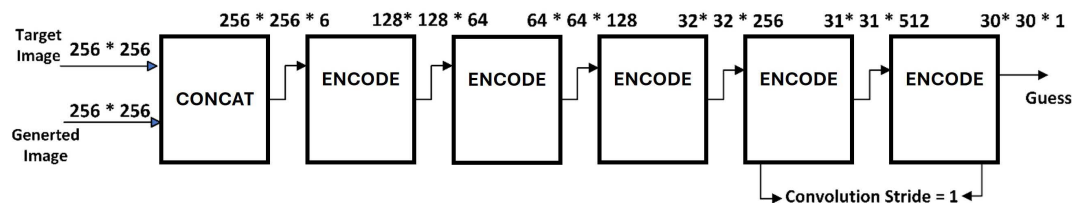


Figure 1. Proposed Discriminator Architecture.

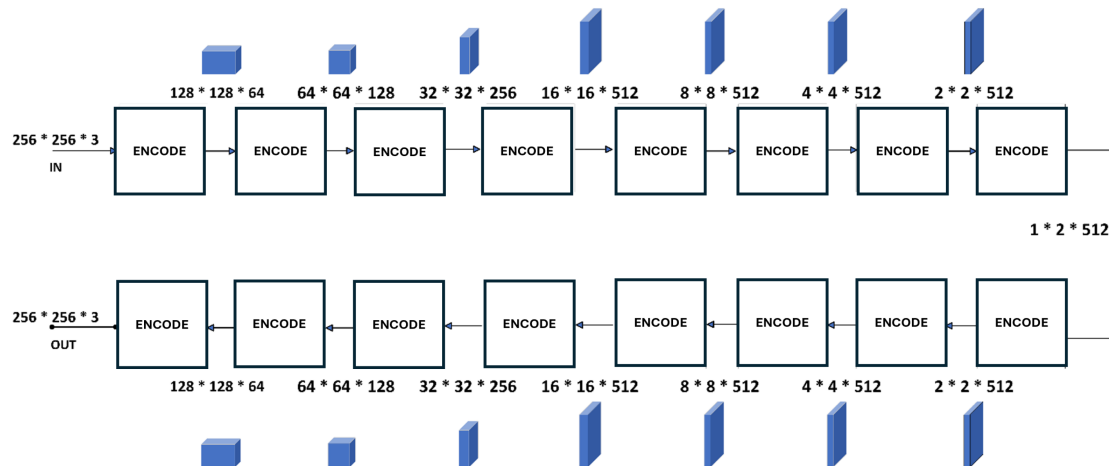


Figure 2. Proposed Generator Architecture.

The skip connection layer connects the encoder to the decoder and passes the convolved features to the decoder which up samples the image from the extracted features passed by the encoder network. At the end an image is produced by the generator which is passed into the discriminator network. The discriminator takes two image pairs as an input; one is the input image and second is the image produced by generator. The discriminator network as shown in Figure 1 is basically an encoder which down samples the images by applying different set of convolutional layers followed by max polling layers on these pairs. Convolutional stride was set to 1 to slightly reduce the size of the image. Finally,

it compares the extracted features and determines whether the image generated by the generator is genuine or not. The discriminator updates its weights based on classification error between the generated image and the original image. The weights of generator are updated based on difference between the generated and target image.

3.4. Conditional Generative Adversarial Network (CGAN)

The goal is to design a model that generates realistic-looking synthetic images using CGAN, therefore choosing the proper parameters is critical. Normally CGAN are hard to train because they suffer from a lot of different issues such as mode collapse. GAN is composed of two CNN based networks that are competing against each other in a min max game; both are trying to win from each other. Initially the generator is not able to produce real looking images and the discriminator easily detects this. In some cases, the generator couldn't produce any real looking images at all whereas sometimes the generator work as an adversary, which fool the discriminator to discriminate between real image or fake image. Setting right set of hyper parameters are crucial in the beginning.

We were able to select the proper parameters after various sets of experimentation, but it was a hit-or-miss method. Because CGAN learns by taking pair images as input, pairs of photos were created to learn the desired mapping. We trained six separate CGAN's to learn six facial expressions i.e., smile, anger, disguise, illumination, stare, wearing glasses.

3.5. Image Pairs Formation for Training and Testing

The 80% set of images containing 107 subjects were separated randomly from the dataset and 20% were kept for testing the performance of the CNN. Neutral face of the same person with smile, anger, disguise, glasses, illumination, and occlusion to learn six facial expressions are shown in Figure 3.



Figure 3. Image Pairs for Training CGAN.

The 20% images set contained 27 subjects with single image per person contains frontal image with neutral pose. Image pairs were formed for testing the six CGAN's individually to generate six facial variations for each person. The image pairs were formed in a way that the single sample per person image was paired with some other subject's facial image variations (i.e. smile, anger, stare, occlusion, illumination, glasses) randomly (as shown in the Figure 4).



Figure 4. AR Dataset- Containing set of paired images for CGAN testing.

3.6. Hyper-parameter Setup

Optimizing the correct set of hyper-parameters is essential from the outset. Through extensive experimentation, we successfully identified the optimal hyper-parameters using a systematic trial-and-error approach. The specifics of these parameters are detailed in Table 1. This table elucidates the configurations for each facial pose, the batch size for paired images, and the generator’s learning direction. Notably, after rigorous testing, the initial learning rate was established at 0.002. Additionally, as indicated in the table, the number of epochs is subject to variation based on further experimentation.

Table 1. Overview of Optimized Hyper-parameters for Facial Pose Generation.

Facial Poses	Batch Size	Generator Direction	Learning Rate	Epochs
Smile	1	A to B	0.0002	650
Anger	1	A to B	0.0002	650
Stare	1	A to B	0.0002	650
Scarf	1	A to B	0.0002	650
Lighting	1	A to B	0.0002	650
Glasses	1	A to B	0.0002	650

4. Dataset

We used the AR dataset [40] to test our model. It contains raw frontal face images containing different expressions. The total number of subjects was 124 and the total number of images was 3300. The total numbers of male subjects were 70 and female subjects were 54. The images were taken in different sessions with different illuminations. We extracted six expressions per person from the database that were smile, anger, stare, illumination, wearing glasses, wearing scarf. We divide our dataset into 80% and 20% set of images randomly. The 80% set of images contained six facial expression per person and the remaining 20% contained one neutral face image per person which was utilized for single sample per person problem. We restricted our experimentation on AR dataset because our method can provide transfer learning. Moreover, our method can produce synthetic samples for any given image.

5. Results and Discussion

In our study, we allocated 20% of the image dataset to test our Conditional Generative Adversarial Network (CGAN), as detailed in Figure 4. This testing resulted in the successful generation of six distinct facial expressions from a single neutral face image, which are showcased in Figure 5.



Figure 5. Generated Six Variations provided Single Neutral Face Image.

During the testing phase, each facial pose image is paired with a neutral face image and then fed into the CGAN as input. Importantly, the facial pose images used for pairing are not specific to any identity or gender, ensuring a broad applicability of the approach. This means that any image, regardless of the depicted individual's identity or gender, can be paired with the neutral face in its neutral pose. The CGAN then leverages this pairing to generate the corresponding expressions on the neutral face image, demonstrating the model's versatility in expression synthesis.

5.1. Training CNN Model on Generated Images

After the six Conditional Generative Adversarial Networks (CGANs) are trained and the synthetic images generated, we employ another network to evaluate these images. For this purpose, we constructed our own Convolutional Neural Network (CNN). All input images are standardized to a resolution of 256×256 pixels, with 3 color channels (RGB). The CNN processes the input image starting with the first layer, which consists of 32 convolutional filters, each of size 3×3 .

A series of convolutional filters of varying sizes (32, 64, 128, and 256) are applied in sequential pairs to extract features at different scales. Following each pair of convolutional filters, a max pooling layer with a 2×2 filter is applied to reduce the spatial dimensions of the feature maps. After the series of convolutional and max pooling layers, a fully connected layer is introduced to integrate the learned features.

The network includes two dense layers, each comprising 4096 nodes, to further process the features. Dropout is implemented as a regularization technique to prevent over-fitting, with a dropout rate designed to prevent the network from becoming too dependent on any one feature. To maintain the stability of the learning process and avoid over-fitting, a momentum of 0.5 is set during training.

5.2. Testing CNN on Original Images Containing Variations

Once we trained our model on the generated samples. The training and validation accuracies and the training and validation losses are shown in Figure 6. We tested the performance of our model on the original images that were present in the database, the original images are shown in Figure 7. These images were separated during the training process of the CGAN and were kept for evaluating the performance of our CGAN generated samples.

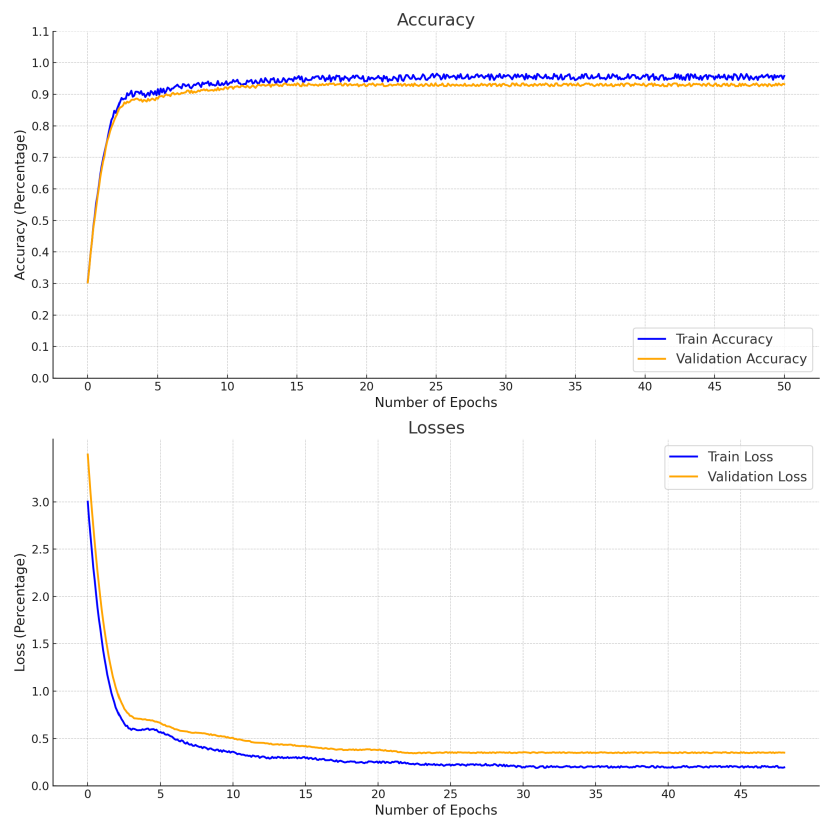


Figure 6. Illustration of Training and Validation Accuracies, along with Training Loss and Validation Loss of the Convolutional Neural Network (CNN).



Figure 7. Illustration of Training and Validation Accuracies, along with Training Loss and Validation Loss of the Convolutional Neural Network (CNN).

5.3. Comparative Analysis of CGAN Generated Samples vs. Single Image per Person using CNN Architecture

We evaluated the effectiveness of our CGAN-generated samples against single images per person using an identical CNN architecture. Our CNN model, trained on single images per person, exhibited promising training results, achieving a 99% accuracy rate. Upon testing the model on six original facial expression variations (smile, anger, stare, occlusion, glasses, and illumination), we attained a testing accuracy of 76%. This result is notably 23% lower than the accuracy of our CNN model trained on CGAN-generated samples.

5.4. Comparison between our Method and State of the Art Method

We benchmarked against state-of-the-art methods, our approach outperformed previous techniques in recognition accuracies, as depicted in Figure 8. Additionally, unlike prior methods that heavily rely on specific datasets, our model demonstrates dataset independence. Once trained on the AR dataset, it can generate synthetic images with variations when provided with only a single image featuring a neutral expression.

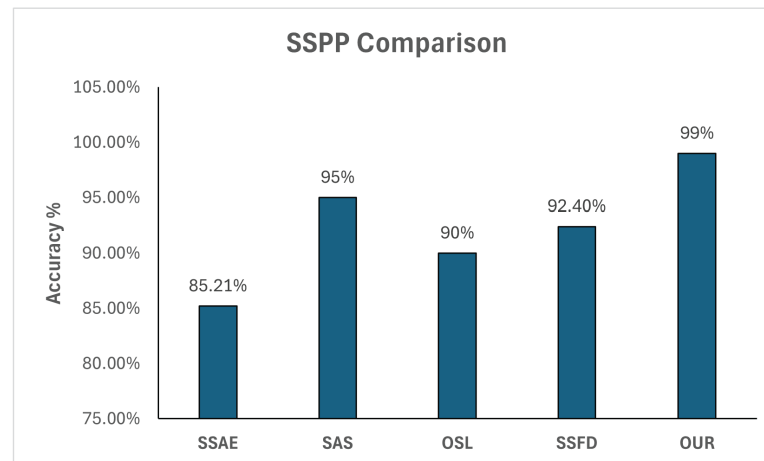


Figure 8. Accuracy Comparison with State-of-the-Art Methods [15,17,18,24].

6. Conclusion

Generative Adversarial Networks have shown some outstanding results when there is lack of data issue associated to real world problem like SSPP and many others. Our research covers the main issues related to the single sample problem. In real world scenarios we are faced by many such scenarios where we must identify a person's image keenly based on only one image which is a very difficult task, the application of SSPP is in important field such as information security, biometrics, access control, driver's license, electronic-passport, identification card and many other automation fields which makes the SSPP problem very significant in the field of facial recognition system. We fine-tuned CGAN to address the issue associated to SSPP. We produced six major facial expression (smile, anger, stare, illumination, glasses, occlusion). Furthermore, our proposed model can also provide transfer learning.

Author Contributions: Conceptualization, M.A.I., W.J.; methodology, M.A.I., W.J. ; software, M.A.I., W.J. ; validation, M.A.I., W.J. and S.K.K.; formal analysis, M.A.I., W.J. and S.K.K.; investigation, S.K.K.; resources, M.A.I., W.J. and S.K.K.; data curation, M.A.I.; writing—original draft preparation, M.A.I.; writing—review and editing, M.A.I.; visualization, M.A.I.; supervision, W.J. and S.K.K.; project administration, W.J.; funding acquisition, S.K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2023RIS-009).

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

References

1. Woubie, A.; Solomon, E.; Attieh, J. Maintaining Privacy in Face Recognition using Federated Learning Method. *IEEE Access* **2024**.
2. Saadabadi, M.S.E.; Malakshan, S.R.; Zafari, A.; Mostofa, M.; Nasrabadi, N.M. A quality aware sample-to-sample comparison for face recognition. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 6129–6138.
3. Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. Face recognition systems: A survey. *Sensors* **2020**, *20*, 342.

4. Benouareth, A. An efficient face recognition approach combining likelihood-based sufficient dimension reduction and LDA. *Multimedia Tools and Applications* **2021**, *80*, 1457–1486.
5. Trevisan de Souza, V.L.; Marques, B.A.D.; Batagelo, H.C.; Gois, J.P. A review on Generative Adversarial Networks for image generation. *Computers & Graphics* **2023**, *114*, 13–25. doi:https://doi.org/10.1016/j.cag.2023.05.010.
6. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE signal processing magazine* **2018**, *35*, 53–65.
7. Hanano, T.; Seo, M.; Chen, Y.W. An improved cgan with self-supervised guidance encoder for generation of high-resolution facial expression images. 2023 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2023, pp. 1–4.
8. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188.
9. Sahan, J.M.; Abbas, E.I.; Abood, Z.M. A facial recognition using a combination of a novel one dimension deep CNN and LDA. *Materials Today: Proceedings* **2023**, *80*, 3594–3599.
10. Vu, H.N.; Nguyen, M.H.; Pham, C. Masked face recognition with convolutional neural networks and local binary patterns. *Applied Intelligence* **2022**, *52*, 5497–5512.
11. Yang, M.; Wang, X.; Zeng, G.; Shen, L. Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. *Pattern recognition* **2017**, *66*, 117–128.
12. Gao, S.; Zhang, Y.; Jia, K.; Lu, J.; Zhang, Y. Single sample face recognition via learning deep supervised autoencoders. *IEEE transactions on information forensics and security* **2015**, *10*, 2108–2118.
13. Abdelmaksoud, M.; Nabil, E.; Farag, I.; Hameed, H.A. A novel neural network method for face recognition with a single sample per person. *IEEE Access* **2020**, *8*, 102212–102221.
14. Gao, S.; Jia, K.; Zhuang, L.; Ma, Y. Neither global nor local: Regularized patch-based representation for single sample per person face recognition. *International Journal of Computer Vision* **2015**, *111*, 365–383.
15. Dang, T.V. Smart attendance system based on improved facial recognition. *Journal of Robotics and Control (JRC)* **2023**, *4*, 46–53.
16. Liu, F.; Tang, J.; Song, Y.; Xiang, X.; Tang, Z. Local structure based sparse representation for face recognition with single sample per person. 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014, pp. 713–717.
17. Hong, S.; Im, W.; Ryu, J.; Yang, H.S. Spp-dan: Deep domain adaptation network for face recognition with single sample per person. 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017, pp. 825–829.
18. Kumar, K.V.; Teja, K.A.; Bhargav, R.T.; Satpute, V.; Naveen, C.; Kamble, V. One-shot face recognition. 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS). IEEE, 2023, pp. 1–6.
19. Yang, S.; Wen, Y.; He, L.; Zhou, M. Sparse Common Feature Representation for Undersampled Face Recognition. *IEEE Internet of Things Journal* **2021**, *8*, 5607–5618. doi:10.1109/JIOT.2020.3031390.
20. Pang, M.; Cheung, Y.M.; Shi, Q.; Li, M. Iterative Dynamic Generic Learning for Face Recognition From a Contaminated Single-Sample Per Person. *IEEE Transactions on Neural Networks and Learning Systems* **2021**, *32*, 1560–1574. doi:10.1109/TNNLS.2020.2985099.
21. Zou, J.; Zhang, Y.; Liu, H.; Ma, L. Monogenic features based single sample face recognition by kernel sparse representation on multiple riemannian manifolds. *Neurocomputing* **2022**, *504*, 82–98.
22. Liu, F.; Wang, F.; Ding, Y.; Yang, S. SOM-based binary coding for single sample face recognition. *Journal of Ambient Intelligence and Humanized Computing* **2022**, *13*, 5861–5871.
23. Petpairote, C.; Madarasmi, S.; Chamnongthai, K. 2d pose-invariant face recognition using single frontal-view face database. *Wireless Personal Communications* **2021**, *118*, 2015–2031.
24. Tsai, T.H.; Chi, P.T. A single-stage face detection and face recognition deep neural network based on feature pyramid and triplet loss. *IET Image Processing* **2022**, *16*, 2148–2156.
25. Neto, P.C.; Sequeira, A.F.; Cardoso, J.S.; Terhörst, P. PIC-Score: Probabilistic Interpretable Comparison Score for Optimal Matching Confidence in Single-and Multi-Biometric Face Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1021–1029.
26. Gui, J.; Sun, Z.; Wen, Y.; Tao, D.; Ye, J. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering* **2021**, *35*, 3313–3332.

27. Singla, K.; Pandey, R.; Ghanekar, U. A review on Single Image Super Resolution techniques using generative adversarial network. *Optik* **2022**, *266*, 169607.
28. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; others. Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.
29. Ning, X.; Gou, D.; Dong, X.; Tian, W.; Yu, L.; Wang, C. Conditional generative adversarial networks based on the principle of homologycontinuity for face aging. *Concurrency and Computation: Practice and Experience* **2022**, *34*, e5792.
30. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning gan for pose-invariant face recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1415–1424.
31. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.