LOW RESOLUTION FACE RECOGNITION USING GENERATIVE ADVERSARIAL NETWORK (GAN)

Howard Martin[1], Suharjito[2]

Computer Science Department, Bina Nusantara University, Jakarta, Indonesia

Abstract

Although face recognition system has achieved a very good performance in the past years, but Low Resolution Face Recognition (LRFR) is still challenging because low resolution image would decrease the accuracy. This research aimed to solved and get the best SR method to solved LRFR problem. YTF dataset used for fine tuning SR methods. While LFW dataset used for fine tuning and evaluating FaceNet model. The images would be increased using Res-Net GAN and RRDB GAN. Then the images would be recognized using FaceNet. The images that had been increased by RRDB GAN reached the highest accuracy 98.96 %.

Keywords: SR; Res-Net GAN; RRDB GAN; Facenet; LRFR

1. Introduction

Face recognition system is a system that has been designed to recognize faces based on the database of faces that has been recognized before [1]. In recent years, we have known a really good improvements in face recognition systems, especially when using a deep learning architeture. One of the state-of-the-art method based on deep learning has been achieved by [2] with an accuracy over 97% on public face datasets such as LFW [3]. Although this algorithm has a good perfomance on good condition of face and pose, these faces generally need to be in high resolution. These designed that has a good performance on High Resolution (HR) image can't be directly used on Low Resolution (LR) image. The surveillance camera has been widely used on public places such as on the streets, office, even in a room (usually placed at the corner). It creates a challenge on face recognition system, where the detected faces would be low in resolution. This research is motivated to solve this challenge.

This work is based on prior work published in [4], but has improved some methods and comparing which method will increase the accuracy for Low Resolution Face Recognition (LRFR). In this work, the resolution of the LR images would be increased and compared using the state-of-the-art method for super resolution by [5] and [6]. There are several method that can be used on super resolution like Bicubic Interpolation [7], Convolutional Neural Network (CNN), Generative Adversarial Network (GAN) [4], etc. Now GAN is the state-of-the-art method that has been used on super resolution comparing with other methods [6]. This work will compare 2 super resolution method, Residual in Residual Dense Block Deep Convolutional GAN (RRDB DCGAN) and Residual Network Deep Convolutional GAN (ResNet DCGAN). The face dataset that will be used on this research is Megaface Challenge 2.

2. Related Works

There are 3 ways to solve the LRFR [7], first method is downsampling the HR images to LR images . The second method is increasing the resolution of LR images using an interpolation methods or super resolution methods [8]. After increasing the resolution of LR images, then the usual face recognition method for HR images can be used to increase the accuracy. The third method is by combining several methods. The most used method is Canonical Correlation Analysis (CCA) [9].

In 2015, [2] used CNN to build a face recognition system. The dataset used in this research are Labelled Faces In The Wild (LFW) [3] and Youtube Faces Database (YTF) [10]. The accuracy result for LFW dataset is 98.95% and YTF dataset is 97.3%. [2] conclude that the error value would keep getting lower if the CNN use more layer.

CNN has reach very good result in face recognition system, especially a face recognition system that has HR images [2], [11], [12]. From there, CNN has been widely used to solved the face recognition problems. CNN has been modified like on [7] who used Multi Resolution CNN. For the pre-processing stage, [7] uses a bicubic interpolation method to decrease the value of cosine similarity. Cosine similarity is a method to measure the similarity between 2 objects. The similarity value ranges from -1 that means exactly opposite and 1 that means exactly similar.

E-mail addresses: howard.martin@binus.ac.id, suharjito@binus.edu,

Table 1 Summary of Related Works

Publication	Problem	Pre-Processing	Method	Evaluation Result
[2]	FR		CNN	LFW: 98.95% YFW: 97.3%
[7]	LRFR	SR with Bicubic Interpolation	Multi-Resolution CNN (MRCNN)	Pie DB: 96.02% Yale DB: 94.63%
[13]	LRFR		CNN with Manifold-based Track Comparison Strategy.	YTF: 80.3%
[14]	LRFR	Extract discriminative features with ResNet Trunk Network	CNN with Deep Coupled ResNet	8x8 pixel: 93.6% 112x96 pixel: 98.7%
[15]	LRFR	Resized image with Bicubic Linear	Facenet	10 x 10: 58.2 % 40 x 40: 98.5 % 160 x 160: 99.2 %
[5]	SR		RRDB GAN	
[6]	SR		ResNet GAN	

[13] also did some research on low resolution face recognition. They did the face recognition from a video using CNN. This research modified the CNN method by adding manifold based track comparison strategy for low resolution video recognition. The dataset used for training are from FaceScrub [16] and MSRA-CFW (dataset of celebrity faces on the web) [17]. YTF dataset is used for testing. This research has reached 80.3% accuracy on YTF dataset.

[14] also researched on low resolution face recognition using Deep Coupled ResNet. The proposed method consists of 1 network trunk and 2 networks branch. Network trunk used for training using 3 different resolution on each face images. Network branch also used for training on the HR and LR images. The objective of this training is to reduce the distance of HR and LR image as small as possible. The parameter on network branch had been optimized using Coupled Mapping. This research used 2 datasets from LFW and SC Face Database [18]. LFW dataset used for testing with 2 different resolution. The 8 x 8 pixels reached 93.6% accuracy and 112 x 96 pixels reached 98.7% accuracy.

[15] had research on face recognition using facenet with several size image. Dataset that had been used on that research was LFW. Every image on the dataset had been resized to 10 x 10, 20 x 20, 40 x 40, 60 x 60, 80 x 80, 100 x 100, dan 160 x 160. Facenet accuracy on 160 x 160 pixel image was 99.2 % and validation rate 97.1 %. But the accuracy gradually decrease with smaller image. Face accuracy on 10 x 10 pixel was 58.2 % and validation rate 1.6 %. The size that would be compared on this research is 40 x 40 pixel. Facenet accuracy on 40 x 40 was 98.5 % and validation rate 88.5 %.

Super resolution method is used to increase the value of cosine similarity. There are several methods of super resolution like bicubic interpolation [19], CNN [20], GAN [9][21], etc. Recently, GAN has achieved an impressive result regarding the Single Image Super Resolution (SISR) problems. Basically, GAN using CNN to do the training phase. GAN architecture is using Deep Convolutional and had been proposed for the first time by Radford on 2015 [22].2 modified GANs that will be used on this work are Residual Network GAN (ResNet GAN) and Residual in Residual Dense Block GAN. ResNet or skip connection is being used on GAN, so that the CNN layers could be used more than 1,000 layers [23] that will cause very minimum error. This two GAN methods has been proposed by [5] and [6]. But those methods had never been applied directly on LRFR problems.

Face recognition method that will be used is FaceNet [24]. FaceNet learns to map face images to an efficient Euclidean space to measure the face similarity based on the distance. FaceNet embeddings as feature vectors can be used by standard methods to solve tasks like face recognition, verification, and clustering [25][26]. For example, k-NN [27] can be used for face recognition by using the embeddings as the feature vectors. All previous works are summarized on Table 1.

3. Theory and Methods

3.1. Low Resolution

Low resolution is one of the main issues in face recognition problems [28]. Those researched and proved that even the state of the art FR methods that reportedly has 99.63% accuracy on LFW [3] and 99.087% on MegaFace Challenge [29] are also degrade significantly on low resolution images. Low resolution images are typically shot and captured using surveillance camera. The image from a great distance captured by the surveillance camera cause the image resolution lower. Those images will lack sufficient visual information for the Deep Learning model to learn the feature representation rather than the HR images. [30] also said that down-sampled images would degrade the accuracy of face recognition.

Generally, there are two ways to solve the LRFR problems. The first way is to decrease the distance between the LR and HR images. The SR methods can be used or even downsample the HR image. The second way is directly used the CNN to learn the feature between LR and HR images. Based on [4], the images lower than 64 x 64 pixels will be classified as low resolution image. This research will use 32 x 32 pixel image.

3.2. MTCNN

MTCNN is a method used for face detection and alignment. CNN has achieved impressive results on this two tasks especially in unconstrained environment [31]. MTCNN used CNN architecture in three stages. The first stage (P-Net) is used for getting the candidate windows throught the shallow CNN. It obtains the candidate facial windows and the box regression vectors. The candidate will be calibrated based on the estimation of bounding box regression vector. Then it used non-maximum supression (NMS) to merge all the highly overlapped faces. The second stage (R-Net) is used for evaluating the windows and rejecting the large number of candidates categorized as non faces. Then it also calibrated with bounding box regression and conducts the NMS. The third stage (O-Net) is used to give an ouput of five facial landmark positions.

3.3. Super Resolution

Super resolution is a method used to increase the resolution of LR image to HR image. Super resolution has widely use to keep the quality of images when the images are strech to a bigger size. Super resolution is not same with only upsizing the image. Super resolution ensures that the upsize images doesn't lose their details and visual appearances. The main goal of super resolution is to find the missing pixel value in the HR image [32].

Super resolution can be divided into two classes: Dynamic resolution (multiple image super resolution) and static resolution (single image super resolution). Dynamic resolution method are based on the assumption that the multiple image are basically geometric transform and misaligned version of each other. Combining and evaluate them together will result more details and fill the missing pixel in HR image. But this method is difficult because the multiple are not always available. So the static resolution is more applicable and will be used for this research. There are many method that can be used for super resolution like bicubic linear, nearest neighbor, generative adversarial network (GAN), etc. But recently GAN has reached a very good performance on super resolution and become the state of the art in super resolution problem.

3.4. Generative Adversarial Network (GAN

GAN is a better method for super resolution because with GAN the generated image looks more real. GAN firstly proposed by Ian Goodfellow through the concept of game theory on 2014 [21]. Basically, GAN consists of 2 neural networks: Generator and Discriminator. Generator (G) used for generating a fake image and Discriminator (D) used for learning what features that make image real. The main principile of GAN is make G and D to compete each other. G

tries to generated an image so that D can not recognize if that image is not a real image. In the other hand, D keep learning on decide whether the input image is real or not. The GAN equation can be seen below.

$$\min_G \max_D V(D,G) = \mathbb{E}_{I^{HR} \sim p-train(I^{HR})}[\log D(I^{HR})] \\ + \mathbb{E}_{I^{LR} \sim pG(I^{LR})}[\log(1 - D(G(I^{LR})))]$$

(1)

4. Proposed Method

1. Dataset

This research will be using 2 face dataset for training and evaluation phase. Youtube Faces Database (YTF) [10] as a training phase dataset on SR neural network. Labelled Faces In The Wild (LFW) [3] will be used as training on FaceNet neural network and also used as evaluation phase. The first dataset is YTF. It is a dataset that designed especially for solving the face recognition problems in video. This dataset contains 3,425 videos of 1,595 subjects. The average video for every subject is 2.15. For each subject, 10 sample images will be used for training the neural network.

The second dataset is LFW. It is a dataset that becomes a standard benchmark for automatic face verification. This dataset contains 13,233 images with 5,749 subjects. It also provides 6,000 pairs of sample data for evaluation phase. Images in this database is also has various challenges like variations of pose, expression and illumination.

YTF and LFW dataset will be prepared by cropping and aligning. Every image on the dataset will be checked if any face is detected on the image. If there is a face detected, the image will be cropped precisely on the face only. So that the background image and other attributes will not affect when face being recognized. The image face also being aligned to straight position. MTCNN method would be used for this task. MTCNN would ensure all the images is still detected as a face. The goal of this process is to get the cropped face and align the face. The sample output of MTCNN shown at Figure 1.

2. Super Resolution

The super resolution will get the input of cropped and aligned images. Fine tuning SR needs Low Resolution (LR) and Ground Truth (GT) from image D. D can be defined as
$D = \{(I_1^{LR}, I_1^{GT}), (I_2^{LR}, I_2^{GT}), (...,...), (I_n^{LR}, I_n^{GT})\}$,
where I is a face image, I^{LR} is the low resolution, I^{GT} is the
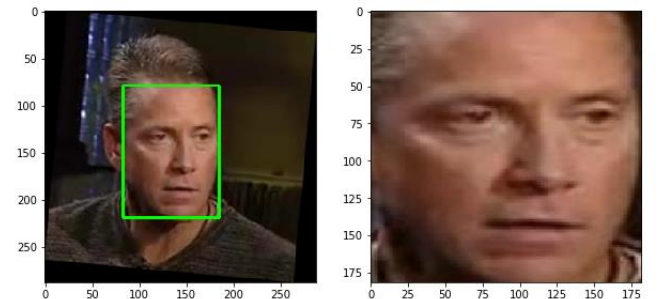

Figure 1 Align and crop face image

image with target resolution, and n is the total image from the

dataset. I^{GT} is an image get from cropping the face image with 128 x 128 pixel. While I^{LR} is resized from I^{GT} to 32 x 32 pixel. The preparation of LR and GT image is shown on Figure 2.
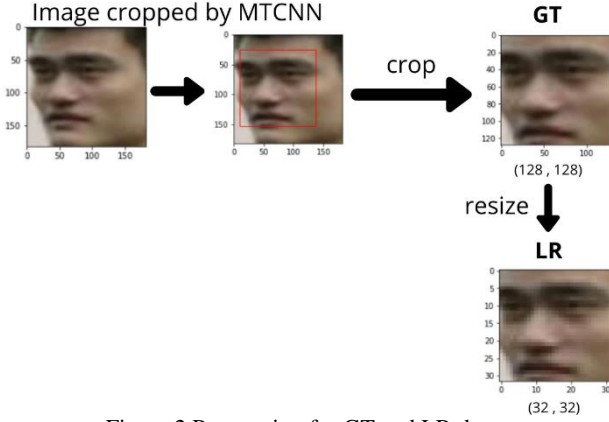

Figure 2 Preparation for GT and LR dataset

The 32 x 32 pixel images will be upsampled using both SR methods, ResNet GAN and RRDB GAN. Those methods will enlarge the images 4x. So the output size of these SR would be 128 x 128. Then the image will be feed and recognize using FaceNet. Both SR methods would be explained further on section 4.3.1 and 4.3.2.

4.3.1. Residual Network GAN (Res-Net GAN)

The ResNet GAN used for this research is also called SRGAN that had been proposed by [6]. SRGAN basically used the ResNet GAN Architecture, but there are several modification on calculate the perceptual loss function.

In this single image super resolution (SISR) task, SRGAN is aiming to predict the HR image and resulting the super-resolved image I^{SR} from the LR image I^{LR}. On the dataset, I^{LR} and I^{GT} are a pair from the dataset, I^{SR} is the predicted image from SRGAN. Based on the basic GAN, the ultimate goal of this method is to train the G network and the D network as the generator and discriminator network. The G network is trained using feed-forward CNN G_{θ_G} where θ_G denotes as the parameter for G. The value of $\theta_G = \{W_{1:L}, b_{1:L}\}$, where W and b denotes as the weight and the bias of L-Layer deep network. The parameter is obtained by optimizing the loss function l^{SR}. From the pre-processing step we get I_i^{LR} where $i = 1, 2, \ldots, N$ and I_i^{HR} where $i = 1, 2, \ldots, N$. So the function:

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{i=1}^{N} l^{SR}\left(G_{\theta_G}(I_i^{LR}), I_i^{HR}\right). \qquad (2)$$

According to the proposed formula of GAN [21] the objective of GAN is to solve equation 1. As explained on the section 3.4, the goal of this equation is for getting G to generate a fake image that can fool D. D is also trained to classify whether the input image is a fake or real. So at the end, the generator can generate an image that can hardly classify by D.

The architecture of generator G consists of 16 residual blocks with the same layout. The layout consists of two convolutional layers with 3x3 kernels, 64 feature maps, batch normalization layers, and ParameticReLU as the activation function. At the end, two sub-pixel convolutional layers are used for increasing the resolution of input image.

To differentiate the fake image and the real image discriminator D needs to be trained. The architecture of D used LeakyReLU with $\alpha = 0.2$ as the activation function. The architecture contains eight convolutional layers with 3x3 filter kernels, increasing factor is 2 from 64 to 512 kernel same as the VGG Network. Strided convolutions are used to maintain the image resolution every time the features is doubled. The result of 512 feature maps are followed by two dense layers and sigmoid activation function as a probability to classify the image.

4.3.2. Residual in Residual Dense Block GAN (RRDB GAN)

Basically RRDB GAN also used residual network GAN, but it modified the network architecture. On the basic ResNet GAN, Generator network architecture contains of convolutional layer, batch normalization (BN) layer, and activation layer. In RRDB GAN it removed all BN layers and add Dense Block (DB) on every blocks.

BN layer is removed from the basic block on ResNet GAN because it has proven in increasing the performance and also reduce the complexity on SR [33] and deblurring [34] tasks. BN layer is used to do the normalization of feature using mean and variance of a current batch during training phase. Meanwhile on the testing phase, BN layer will normalize the feature using estimated mean and variance of the whole training dataset. In case when the training and testing dataset differs a lot, then BN layer will reduce the performance and limit the generalization ability. So [5] removed BN layer to produce a stable and consistent performance during training.

In CNN, more layer and connection will always result better performance [35][36]. This RRDB network use dense block on the main path that will cause the network capacity becomes better with the dense connection. Residual scaling and smaller initialization were also used to improve the architecture. Scaling down the residual with β where $0 < \beta < 1$ before adding to the main path will stable the process. Residual architecture also train easier when the initial parameter variance is smaller.

There is also improvement on the Discriminator based on Relativistic GAN. The standard Discriminator (D) evaluates if images x is naturally real or fake. But relativistic discriminator will evaluates if image x_r is relatively more realistic than the fake x_f. Relativistic Discriminator D_{Ra} can be calculated as

$$D_{Ra}(x_r, x_f) = \sigma(C(x_{rf}) - \mathbb{E}_{x_f}[C(x_f)]) \qquad (3)$$

Where σ is a sigmoid function, C is the discriminator output, and $\mathbb{E}_{x_{fake}}$ is the average off all the fake images in a single batch.

3. Facenet

FaceNet is a method proposed by [24]. It uses a deep

CNN architecture with triplet loss to get an embedding that can be used for face clustering, verification, and recognition. FaceNet provides a unified embedding $f(I)$ that maps face I into a d – dimensional Eucledian space, where $f(I) \in \mathbb{R}^d$. This triplet loss method is inspired by nearest-neighbor classification. Triplet loss is a loss function that will ensure that an image of a person face I_i^a (anchor) is closer to all other image I_i^p (positive) of the same person than to all other image I_i^n (negative) of other person. The triplet loss function is shown at Figure 3.
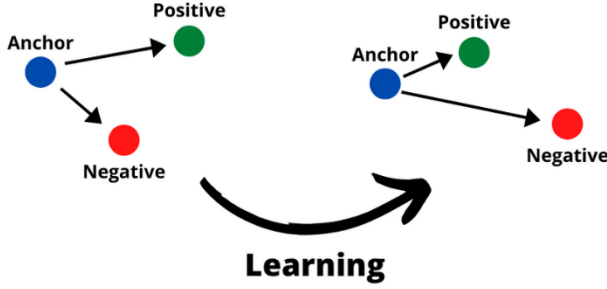

Figure 3: Triplet loss will minimize the distance between anchor and positive (same person) and maximize the distance between anchor and negative (different person)

The triplet loss function L can be defined as

$$L = \sum_{i=1}^{N} \left[\left\| f(I_i^a) - f(I_i^p) \right\|_2^2 - \left\| f(I_i^a) - f(I_i^n) \right\|_2^2 + \alpha \right]_+,$$

(4)

Where N is the amount of image in a set of all possible triplets pair on the training set and α is a difference margin between the anchor positive and anchor negative.

FaceNet trains the CNN using Stochastic Gradient Descent (SGD) with backprop and AdaGrad. The initial learning rate is 0.05, the α is 0.2 and ReLU is the activation function.

The LFW dataset from SR is used on the training and evaluation phase. Every identity will be splitted into training set, validation set and testing set. The portion of the dataset is 90% for training, 5% for validation, and 5% for testing. The minimum image for each identity is 4 images. 2 for training, 1 for validation, and 1 for testing. Validation is used to prevent the training process to get underfitting or overfitting.

During training and testing $L2$ – distance threshold will be set on 1.242 based on [24] experiment. The squared $L2$ of $D(I_i, I_j)$ is used to differentiate if those pair image is same or different. All pair faces (i, j) of the same person is denoted as P_{same} and all pair faces (i, j) of different person is denoted as P_{diff}. A set of true accept identities are calculate with

$$TA(d) = \{(i,j) \in P_{same}, \quad D(i,j) \leq d\}.$$

It will create a set of face pair (i, j) that correctly classified as a same person with threshold d. Then

$$FA(d) = \{(i,j) \in P_{diff}, D(i,j) \leq d\}$$

will create a set of face pair (i, j) that correctly classified as a different person with threshold d.

The Validation rate $(Val(d))$ will be calculated with $Val(d) = \frac{|TA(d)|}{|P_{same}|}$, and the False accept rate $(Far(d))$ will be calculated with $Far(d) = \frac{|FA(d)|}{|P_{diff}|}$.

5. Results and Discussion

The training result for FaceNet using the dataset that had been increased by Res-Net GAN can be seen on Figure 4. After using the pre-trained weight, the accuracy on epoch 0 reached by Res-Net GAN was 75.5%. The accuracy increased significantly on the first 20 epoch. The accuracy reached on epoch 20 was over than 90 %. But the accuracy increased slowly on the next epoch. The accuracy started to converge approximately on epoch 70. On epoch 70 the accuracy had reached 97 %. Until epoch 100, the accuracy that had been reached by FaceNet using Res-Net GAN was 97.02 %. On the other hand, the validation accuracy on epoch 0 reached 92.3%. The validation accuracy increased up to 96% on epoch 50. After that the accuracy still increased slowly and converge approximately on epoch 60. Until epoch 100, the accuracy using the validation set by FaceNet using Res-Net GAN was 98.03 %.

The training loss that had been reach on epoch 0 was 9.546. Loss value decreased significantly on epoch 20. Loss value on that epoch was approximately at 2.5. The training loss kept decreasing until 1.5 at epoch 50. Training loss started to converge at epoch 60. Until epoch 100, the training loss that had been reached by FaceNet using Res-Net GAN was 0.935. The validation loss on epoch 0 was 1.68. Validation loss kept decreasing until 0.5 on epoch 20. Validation loss started to converge approximately at epoch 80. The validation loss reached on epoch 100 was 0.237.
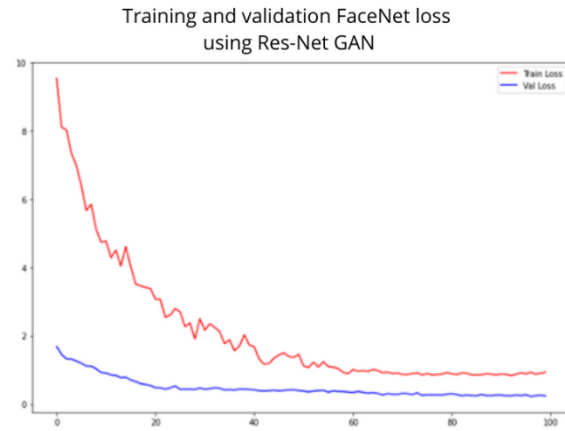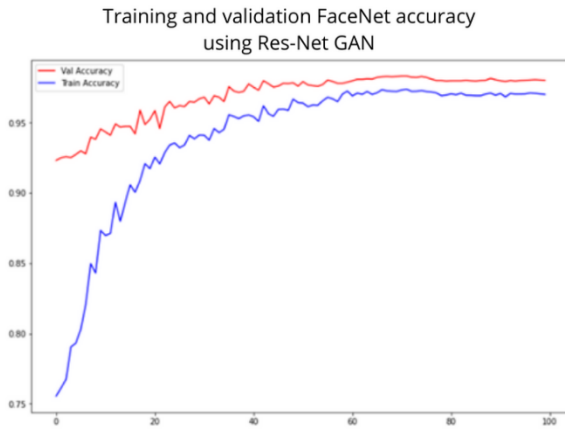
Figure 4: Accuracy and loss on facenet using Res-Net GAN

Figure 5: Accuracy and loss on facenet using RRDB GAN

The training result for FaceNet using the dataset that had been increased by RRDB GAN can be seen on Figure 5. After using the pre-trained weight, the accuracy on epoch 0 reached by Res-Net GAN was 76.4%. The accuracy increased significantly on the first 20 epoch. The accuracy reached on epoch 20 was approximately on 93 %. On the next epoch, the accuracy kept increased slowly until epoch 70. The accuracy on epoch 70 was approximately on 97%. The accuracy started to converge approximately on epoch 85. Until epoch 100, the accuracy that had been reached by FaceNet using RRDB GAN was 98.4 %. The validation accuracy on epoch 0 reached 92.5%. The validation accuracy increased until 95.5% on epoch 20. Validation accuracy kept increased slowly until 97% on epoch 70. It started to converge on epoch 80. Until epoch 100, the accuracy using the validation set by FaceNet using RRDB GAN was 98.4 %.

The training loss that had been reach on epoch 0 was 9.46. Loss value decreased significantly until epoch 30. Loss value on that epoch was approximately at 2.0. The training loss kept decreasing until epoch 60. Training loss on epoch 60 was approximately at 1.0. Until epoch 100, the training loss that had been reached by FaceNet using Res-Net GAN was 0.92. The validation loss on epoch 0 was 1.46. Validation loss kept decreasing until 0.3 on epoch 30. Validation loss started to converge approximately at epoch 80. Validation loss on that epoch was 0.27. The validation loss reached on epoch 100 was 0.244.
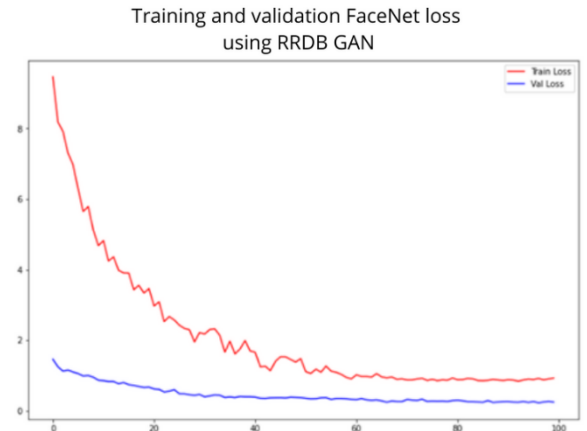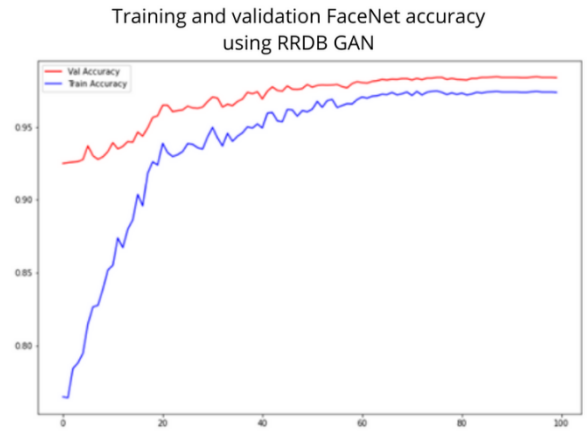
After training the facenet model using training and validation set, the model was tested using the testing set. The testing comparison result can be seen on Table 2. The accuracy reached by FaceNet using Res-Net GAN was 95.53 % and validation rate was 94.667 %. This result is better than [15] on accuracy by 0.03% and on validation rate by 6.167%. The accuracy on Facenet using RRDB GAN reached 98.96 % and validation rate was 96.757 %. This result is the best result on this research. It is better than [15] on accuracy by 0.46% and on validation rate by 8.257%.

6. Conclusion

This researched proposed a model to solve the LRFR problem. This research proposed SR methods using Res-Net GAN and RRDB GAN before recognition method using FaceNet. This model was evaluated using LFW dataset. The result showed that using SR before face recognition can increase the accuracy. It also showed that using RRDB GAN to increase the low resolution image gave the best result. For further research, it could be added some features on face recognition model like gender and age category. It would narrow down the possibility for facenet to false predict.

Table 2 Comparison result on FaceNet

SR Methods	Face Recognition	Size	FaceNet Evaluation	
			Accuracy	VAL

Res-Net GAN	FaceNet	32 x 32	98.53 %	94.67%
RRDB GAN	**FaceNet**	**32 x 32**	**98.96 %**	**96.76%**
-	[15]	40 x 40	98.50 %	88.50%

Conflict of interest

The authors declare that there is no conflict of interest in this paper.

Abbreviations

GAN: Generative Adversarial Network;

Acknowledgements

Not Applicable

Author's contributions

HM designed and experimented this study in consultation with S. S supervised the process, results, and design of this study. HM wrote this report and being fully supervised by S. Both authors have contributed in this study and have approved the submitted version.

Availability of data and materials

This study used the face datasets from [3] and [37]. It is available by downloading the dataset from the website.

Competing Interests

Not Applicable

Author Details

The authors declare that there is no conflict of interest in this paper.

References

[1] S. Singh and S. V. A. V. Prasad, "Techniques and challenges of face recognition: A critical review," *Procedia Comput. Sci.*, vol. 143, pp. 536–543, 2018.

[2] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," *BMVC*, no. Section 3, pp. 41.1-41.12, 2015.

[3] G. B. Huang, M. Mattar, T. Berg, and E. Learned-miller, "Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments," *Tech. rep.*, 2008.

[4] P. Li, L. Prieto, D. Mery, and P. Flynn, "On low-resolution face recognition in the wild: Comparisons and new techniques," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, pp. 2000–2012, 2018.

[5] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2019.

[6] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 105–114, 2017.

[7] T. Bao, C. Ding, S. Karmoshi, and M. Zhu, "Video-based face recognition via convolutional neural networks," *Second Int. Work. Pattern Recognit.*, vol. 10443, p. 104430I, 2017.

[8] J. W. Hwang and H. S. Lee, "Adaptive Image Interpolation Based on Local Gradient Features," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 359–362, 2004.

[9] W. S. Lai, J. Bin Huang, N. Ahuja, and M. H. Yang, "A Fully Progressive Approach to Single-Image Super-Resolution Supplemental Material," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5835–5843, 2017.

[10] L. Wolf, T. Hassner, and I. Maoz, "Face Recognition in Unconstrained Videos with Matched Background Similarity."

[11] H. M. Chang, H. Seo, and S. Bum, "A face recognition system based on convolution neural," *Soft Comput.*, 2016.

[12] X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," pp. 2–6, 2015.

[13] C. Herrmann, D. Willersinn, and J. Beyerer, "Low-resolution Convolutional Neural Networks for video face recognition," *2016 13th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2016*, no. August, pp. 221–227, 2016.

[14] Z. Lu, X. Jiang, and A. Kot, "Deep Coupled ResNet for Low-Resolution Face Recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 526–530, 2018.

[15] M. R. Golla and P. Sharma, *Performance Evaluation of Facenet on Low Resolution Face Images*. Springer International Publishing, 2019.

[16] S. Ng, Hong-Wei, Winkler, "A DATA-DRIVEN APPROACH TO CLEANING LARGE FACE DATASETS Hong-Wei Ng and Stefan Winkler Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign , Singapore," *2014 IEEE Int. Conf. Image Process.*, pp. 343–347, 2014.

[17] X. Zhang, L. Zhang, S. Member, and X. Wang, "Finding Celebrities in Billions of Web Images," vol. 14, no. 4, pp. 995–1007, 2012.

[18] M. Grgic, K. Delac, and S. Grgic, "SCface – surveillance cameras face database," no. October 2009, pp. 863–879, 2011.

[19] V. K. Tanwar, "Privacy preserving image scaling using 2D bicubic interpolation over the cloud," *2018 IEEE Int. Conf. Syst. Man, Cybern.*, pp. 2073–2078, 2018.

[20] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, and Q. Liao, "Deep Learning for Single Image Super-Resolution :," no. 1, pp. 1–16.

[21] I. J. Goodfellow et al., "Generative Adversarial Networks," pp. 2672–2680, 2014.

[22] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," pp. 1–16, 2015.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.

[24] F. Schroff and J. Philbin, "FaceNet : A Unified Embedding for Face Recognition and Clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[25] S. Karamizadeh and S. M. Abdullah, "Race classification using gaussian-based weight K-nn algorithm for face recognition," vol. 6, no. M, 2018.

[26] H. S. Dadi, G. Krishna, and M. Pillutla, "Improved Face Recognition Rate Using HOG Features and SVM Classifier," vol. 11, no. 4, pp. 34–44, 2016.

[27] C. Eyupoglu, "Implementation of Color Face Recognition Using PCA and k -NN Classifier," pp. 199–202, 2016.

[28] S. Cheng, Zhiyi and Zhu, Xiatian and Gong, "Low-Resolution Face Recognition," vol. 1, Springer International Publishing, 2019, pp. 605–621.

[29] I. K. Steven and M. S. Daniel, "The MegaFace Benchmark : 1 Million Faces for Recognition at Scale," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4873–4882, 2016.

[30] P. E. I. Li and N. Dame, "Face Recognition in Low Quality Images : A Survey," vol. 1, no. 1, 2019.

[31] K. Zhang, Z. Zhang, Z. Li, S. Member, Y. Qiao, and S. Member, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, 2016.

[32] R. A. A. Al-falluji, A. A. Youssif, and S. K. Guirguis, "Single Image Super Resolution Algorithms : A Survey and Evaluation," *Int J Adv Res Comput Eng Technol*, vol. 6, no. 9, pp. 1445–1451, 2017.

[33] B. Lim and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution."

[34] S. Nah, T. H. Kim, and K. M. Lee, "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring."

[35] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual Dense Network for Image Super-Resolution."

[36] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," pp. 15–17.

[37] L. Wolf, T. Hassner, and I. Maoz, "Face Recognition in Unconstrained Videos with Matched Background Similarity," *CVPR 2011*, pp. 529–534, 2011.