IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.16, No.1, January 2022, pp. 1~10
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.58327

Face Image Generation and Enhancement Using Conditional Generative Adversarial Network

Ainil Mardiah*[1], Sri Hartati[2], Agus Sihabuddin[3]
[1]Master Program in Computer Science, FMIPA UGM, Yogyakarta, Indonesia
[2,3]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
e-mail: ***[1]ainil.mardiah@mail.ugm.ac.id**, [2]shartati@ugm.ac.id, [3]a_sihabudin@ugm.ac.id

Abstrak

Akurasi dan kecepatan single image super-resolution menggunakan convolutional neural network sering menjadi masalah dalam perbaikan detail tekstur yang lebih halus ketika menggunakan faktor peningkatan yang besar. Beberapa penelitian terbaru terfokus pada mean square error minimal, sehingga menghasilkan peak signal to noise ratio yang tinggi. Umumnya, walaupun peak signal to noise ratio memiliki nilai yang tinggi, namun citra output kurang detail. Hal ini menunjukkan penentuan super-resolution belum optimal. Conditional Generative Adversarial Network berbasis Boundary Equilibrium Generative Adversarial Network, dengan menggabungkan Mean Square Error Loss dan GAN Loss sebagai fungsi loss untuk mengoptimalkan model super-resolution dan menghasilkan citra super resolution. Disamping itu, jaringan generator dirancang dengan arsitektur skip connection untuk meningkatkan kecepatan konvergensi dan memperkuat penyebaran fitur.

Parameter nilai kualitas citra yang digunakan pada penelitian ini adalah Peak Signal to Noise Ratio (PSNR) dan Structral Similarity Index (SSIM). Hasil penelitian menunjukkan nilai kualitas citra tertinggi menggunakan validation dataset adalah 26,55 dB untuk nilai PSNR dan 0,93 untuk nilai SSIM. Nilai kualitas citra tertinggi menggunakan testing dataset adalah 24,56 dB untuk nilai PSNR dan 0,91 untuk nilai SSIM.

Kata kunci*— Conditional GAN, Boundary Equilibrium, Single Image Super-Resolution*

Abstract

The accuracy and speed of a single image super-resolution using a convolutional neural network is often a problem in improving finer texture details when using large enhancement factors. Some recent studies have focused on minimal mean square error, resulting in a high peak signal to noise ratio. Generally, although the peak signal to noise ratio has a high value, the output image is less detailed. This shows that the determination of super-resolution is not optimal. Conditional Generative Adversarial Network based on Boundary Equilibrium Generative Adversarial Network, by combining Mean Square Error Loss and GAN Loss as a loss function to optimize the super-resolution model and produce super-resolution images. Also, the generator network is designed with skip connection architecture to increase convergence speed and strengthen feature distribution.

Image quality value parameters used in this study are Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The results showed the highest image quality values using dataset validation were 26.55 for PSNR values and 0.93 for SSIM values. The highest image quality values using the testing dataset are 24.56 for the PSNR value and 0.91 for the SSIM value.

Keywords*—Conditional GAN, Boundary Equilibrium, Single Image Super-Resolution*

Received August 1st,2020; Revised January 31th, 2022; Accepted January 31th, 2022

1. INTRODUCTION

Image is a complex and high dimension that is difficult to make a good model. A model that can explain how data is generated from data distribution is known as generative model. Building a good natural image generative model is a fundamental problem in computer vision [1]. Generative model allows machine learning to work with multi –modal output. The example of some generative models that requires a good sample generation is Single Image Super Resolution (SISR). This method aims to use low resolution (LR) images and unite them into an equivalent high resolution (HR) images [2].

SISR method is used to generated super resolution image [3]. Super resolution image, which means that the object in the image is sharp and detail, has many application in remote sensoring, medical diagnostic, intelligent observation and others. A high resolution (HR) image can provide more detail than its low resolution pair and this detail is important in many applications. In the most cases, face image appear in the form of LR due to limitations in producing samples, storage and dissemination of HR images, such as taking pictures of LR faces with a CCTV camera. So to get more details, then a series of LR images must concluded to be an HR image. This technique is called Super Resolution (SR) [4].

The implementation of SR method needs a generative model that can generate a HR image based on the corresponding LR image. One of generative model that is widely used today is Generative Adversarial Network (GAN) [5]. GAN has recently used as an alternative framework to train the generative model to avoid the difficulty of estimating many probabilistic calculation that difficult to solve [6]. In the implementation, GAN method is combined with some additional techniques to produce better result.

In this study, the LR image is used as a condition to produce (with 4x upscaling factor) an HR image that can represent the original image. The application of additional condition on the GAN network was not only able to produce image with more specific detail, but also can be used to produce a prediction. the addition of the additional condition is known as Conditional Generative Adversarial Network (CGAN) [6][7]. Even though GAN can produce impressive image, but GAN still faces some unresolved problems. Difficulties in the training stage are one example of problems faced by GAN, so choosing the right hyper parameters is important [8]. The optimization parameters of the generator can produce a good sample image and discriminator cannot distinguish between the sample produced bby the generator from the original sample. This problem causes damage to the balance of the generator and discriminator, so generator and discriminator do not reach the optimal level. Some research that examines problems encountered during the training process are Energy based Generative Adversarial Network [9] and BEGAN: Boundary Equilibrium Generative Adversarial Network [8].

The BoundaryEquilibrium Generative Adversarial Network method is able to balance the network of generator and discriminator during the training process, so as to produce image with better quality. Therefore, the Equilibrium algorithm is used to maintain the balance of generator and discriminator during the training process.

2. METHODS

The purpose of Super-Resolution (SR) is mapping LR I^{LR} images into HR I^{SR} images. I^{LR} is the original image that has been downsampled. Study of Huang et al. (Huang et al., 2018) use the Conditional Generative Adversarial Network to produce I^{SR} and show I^{SR} result that represents I^{HR}. In this study, aside from being used as an input image, I^{LR} it is also used as a condition to produce I^{SR}. GAN is very easy to experience capital collapse [10][9], Boundary Equilibrium GAN (BEGAN) [8] is used to balance the convergence between the generator

network and the discriminator during the training process [2]. BEGAN uses a loss function derived from Wasserstein distance [11], the main purpose of BEGAN is to optimize Wasserstein distance between loss distributions. BEGAN is used to balance the convergence of the generator and discriminator networks, calculate the optimization value of the GAN model, and solve the model collapse problem [12].

2.1 Conditional Generative Adversarial Network

Generative Adversarial Network (GAN) was introduced by Goodfellow et al. [5] in 2014, the artificial intelligence algorithm used in unsupervised machine learning. This technique produces images or photos that look original to human vision because they have many realistic characteristics. The basic idea behind GAN is to train two networks, namely the G(z) generator network that produces a face image and the D(x) discriminator network that attempts to distinguish the image generated by a generator or fake network from the original image [12]. One of GAN that is widely used today is the Conditional Generative Adversarial Network. This type of GAN uses additional requirements added in the generator network to produce an HR image. The architecture of the Conditional GAN generator and discriminator network is shown in Figure 1:

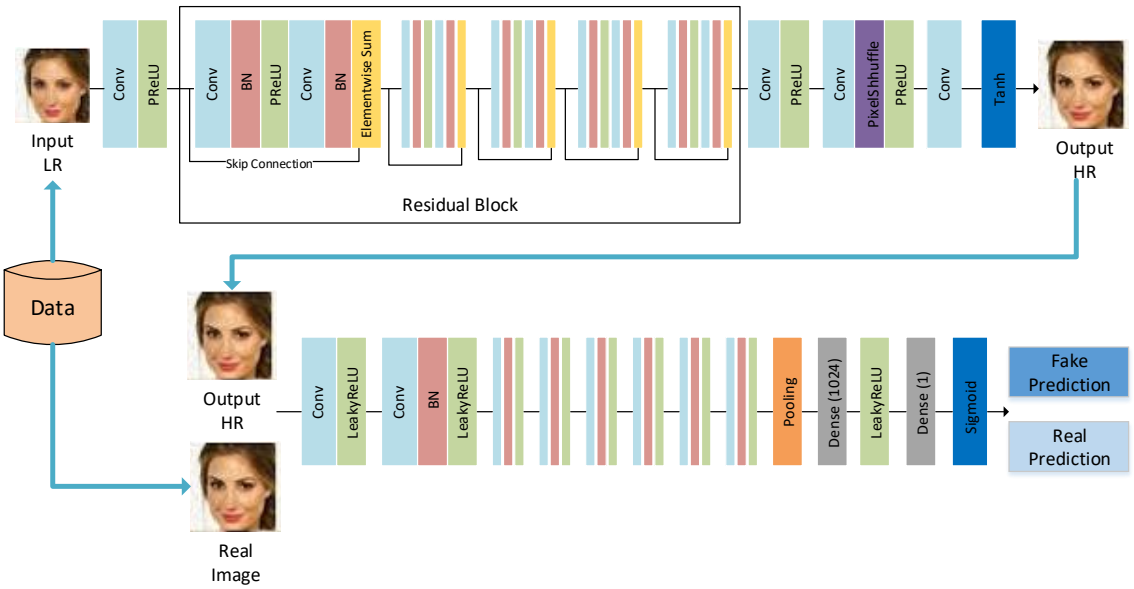

Figure 1 Generator and Discriminator Network Architecture of Conditional GAN

2. 1.1 Generator Network

As Figure 1 shows, the implementation of the generator network architecture adapts the structure of the Residual Network (ResNet) [3]. Residual networks are designed in the form of blocks, where each block has two convolution layers with 3x3 kernel size, two Batch Normalization, and PReLU activation functions. Skip connection is used in every residual block stack. The generator network is a network that reconstructs the LR I^{LR} image into an HR I^{HR} image, to increase the resolution of I^{LR}, the generator network using the Upsampling Function. The upsampling function used in this study is the sub-pixel convolution layer introduced by Shi et al. [13].

2. 1.2 Discriminator Network

Discriminator network functions to discriminate input images I^{SR} and original images I^{HR}. The discriminator network architecture is shown in Figure 1, adapting the Deep Convolutional GAN architecture introduced by Radford, Metz, & Chintala [14] and using the

LeakyReLU activation function (α = 0.2). The discriminator network has eight convolution layers with a 3x3 kernel size and an upsampling factor of 4 from 64 kernels to 512 kernels. The 512 feature results are used as input for two dense layers and a sigmoid activation function that will produce sample classification probability.

Generator and discriminator network use parameter specified ini SRGAN paper. The code of generator and discriminator network is available on GitHub.

2. 2 Boundary Equilibrium

GAN is a minimax game where opponents wil try to weaken others, this makes GAN more difficult to convert than other deep learning models. In the generative model, the reconstructed image will be matched with the original image. BEGAN is designed based on the compability or similarity distribution of reconstruction and original image loss.

Similar to Conditional GAN, discriminator network in BEGAN is also designed to calculate the loss model value. The purpose of discriminator network is to minimize the difference between reconstruction image loss and the original image. BEGAN uses Wasserstein distance to measure the difference between reconstruction image loss and the original image. Wassertein distance calculates the transformasion from one distribution to another.

Overfitting can occur in the generator or discriminator if there is no balance in training process, or in the worst case collapse mode can occur during training process. The main idea of BEGAN is having a new loss function using auto encoder as discriminator, where the loss is derived from wassertein distance (to solve collapse mode problem) between reconstruction of the original image loss and the generated image. Hyper parameter of gamma is added using weight parameter of k to provide power for discriminator network to control the desired differences. The code for boundary equilibrium is available on GitHub.

2. 3 Loss Functions

GAN is a method that studies the distribution of *x* data to produce fake data in the *G(x)* generator network and distinguishes the data produced according to the original data or not in the discriminator *D*. In this study, the generator network not only studies the distribution of images but also studies mapping the LR image into an HR image. The objective functions used in this study are adversarial loss and Mean Square Error (MSE) or also called L_2 norm loss. L_2 norm loss is used to calculate the error (loss) between the resulting image and the original image, as shown in equation (1).

$$I_G = G(I_{LR};\theta_G) \tag{1}$$

$$\mathcal{L}_{L2} = (I_G - I_{HR})^2 \tag{2}$$

θ_G is a parameter used by the generator network to produce I^{SR} images. The objective function of the discriminator network is shown in equation (5).

$$\mathcal{L}_{HR} = D(I_{HR};\theta_D) \tag{3}$$

$$\mathcal{L}_{SR} = D(I_G;\theta_D) \tag{4}$$

$\mathcal{L}_{HR}$ is discriminator loss for original images and $\mathcal{L}_{SR}$ is the discriminator for low-resolution images *LR*. θ_D is parameter that used by discriminator network to discriminate reconstruction image and original image.

$$\mathcal{L}_D = \mathcal{L}_{HR} - \mathcal{L}_{SR} \tag{5}$$

Equations (3) and (4) are adopted from BEGAN [8], the core of the BEGAN algorithm is feedback control to maintain the overall balance of the training process. The BEGAN algorithm is characterized by the addition of the parameter γ and weight k in the discriminator network. For the value of $\gamma = 0.7$, $\lambda_k = 0.001$ in the study.

$$\begin{cases} \mathcal{L}_D = \mathcal{L}_{HR} - k_t \times \mathcal{L}_{SR} \\ k_{t+1} = k_t + \lambda_k(\gamma \times \mathcal{L}_{HR} - \mathcal{L}_{SR}) \end{cases} \quad (6)$$

The loss function of the generator in this study uses L_2 norm loss and adversarial loss (shown in equation (7)). Thus, the loss generator function is shown in equation (8), where the adversarial weight used is 0.05 and the L_2 norm loss weight is 0.95.

$$\mathcal{L}_{GAN} = 1 - \mathcal{L}_{SR} \quad (7)$$

$$\mathcal{L}_G = W_{GAN} \times \mathcal{L}_{GAN} + W_{L2} \times \mathcal{L}_{L2} \quad (8)$$

3. RESULTS AND DISCUSSION

This study uses CelebA images as a dataset, totaling 202.599 images. This dataset is then divided into training, validation, and testing data. CelebA dataset has a large number of images, the experiment is divided into 5 batches, they are A1 with 10% of the total dataset, A2 with 20% of the total dataset, A3 with 50% of the total dataset, A4 with the number data 70% of the total dataset and A5 uses 100% of the total dataset.

The results are shown into qualitative and quantitative result. Qualitative results are the comparison of generic images with the original HR images shown in Figure 2 and the comparison of more clearly detail shown in Figure 3. Each rows in Figure 3 compared different detail. The first and the last row demonstrate detail of eye with glasses and without glasses. The middle row demonstrates detail of edge information. Comparison in Figure 3 show that our CGAN model can generate high frequency information and reconstruct tiny detail features.

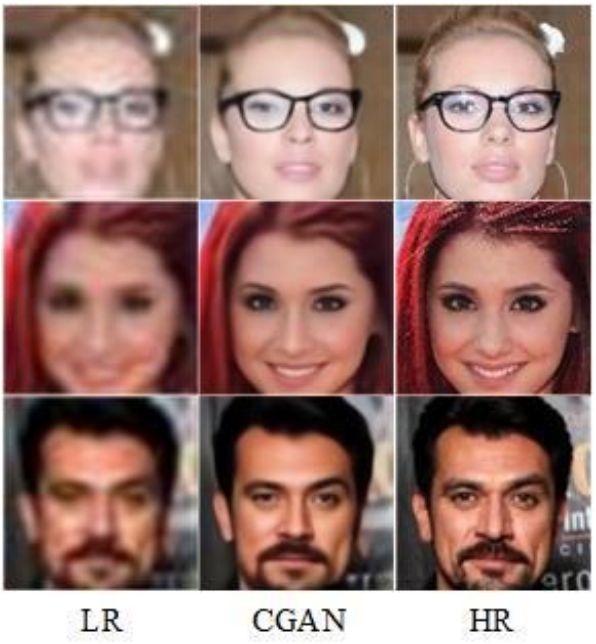

Figure 2 Qualitative Comparison with CelebA Dataset (4x upscaling factor)

Figure 3 Comparison of Image's Detail

Quantitative results are the comparison of PSNR and SSIM values for each experiments shown in Table 1 using validation dataset and Table 2 using testing dataset. Quantitative results is also used for the comparison of PSNR and SSIM values with a previous study shown in Table 3.

Table 1 Quantitative Comparison Using Validation Dataset

Experiment	Number of Images	Quantitave Value of Image	
		PSNR	SSIM
A1	1.013	26,036	0,923
A2	2.026	26,346	0,927
A3	5.065	26,553	0,931
A4	7.091	26,646	0,931
A5	10.130	26,634	0,931

Table 2 Quantitative Comparison Using Testing Dataset

Experiment	Number of Images	Quantitave Value of Image	
		PSNR	SSIM
A1	1.013	24,564	0,909
A2	2.026	24,507	0,906
A3	5.065	24,50	0,907
A4	7.091	24,501	0,907
A5	10.130	24,507	0,906

Table 3 Comparison of PSNR and SSIM Value with The Previous Study

Number of Image	Size of Input Image	Quantitave Value of Image	
		PSNR	SSIM
10.000	128 × 128	32,66	0,863
10.130	32 × 32	24,51	0,906

Features contained in input image can be lost when using multi-layer convolutional nework (CNN). Model also can not reach the performance of the model with skip-connection, although increasing training iteration. It means that skip-connection can keep useful information, even though deep convolutional network can not recover in the next layer [12].

This study uses skip-connection to improve the performance of the CGAN model and accelerate conversions during the training process. The training curve is shown in the graph of generator and discriminator network loss using skip-connection during the training process. The X coordinate shows the iteration during the training phase (with the number of epochs being 10) and the Y coordinate shows the value of the generator and discriminator network loss functions. The training process uses the same number of epoch in each experiment. Figure 4 and 5 shows the training curve using 10% and 20% of the total dataset. Figure 6 and 7 shows the training curve using 50% and 70% of the total dataset. Figure 8 shows the training curve using 100% of the total dataset.

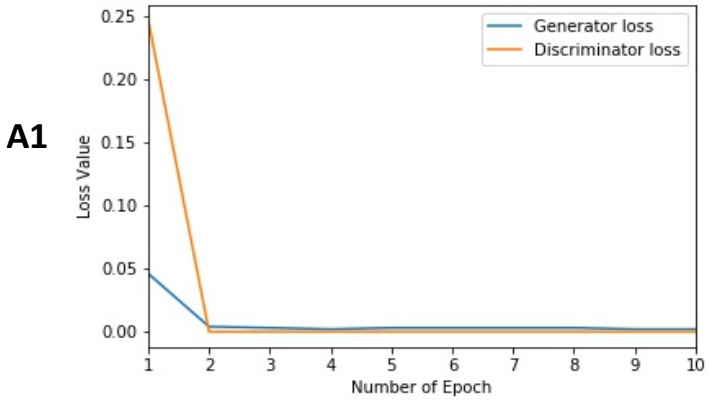

Figure 4 The Training Curve of A1 Experiment

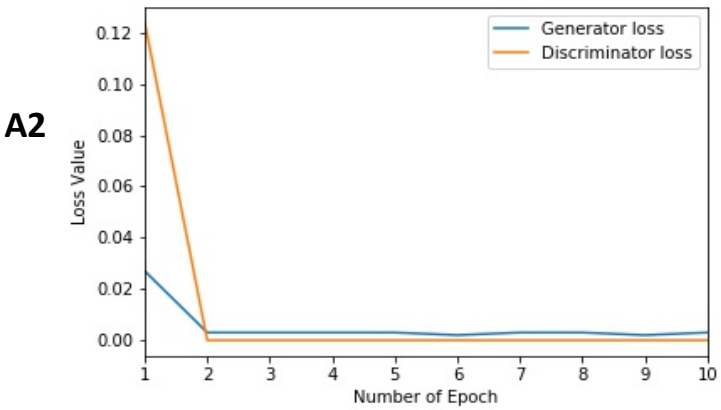

Figure 5 The Training Curve of A2 Experiment

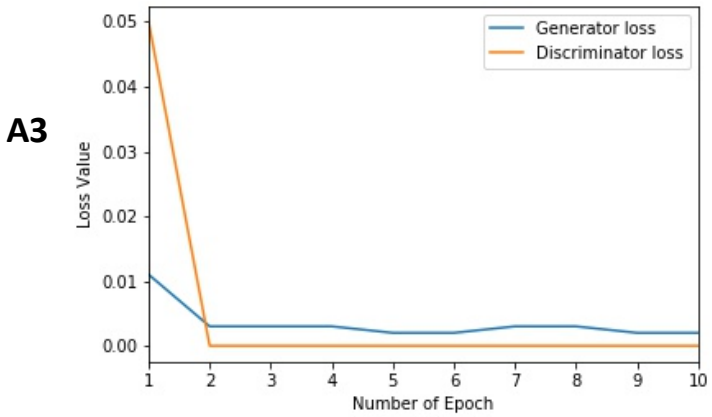

Figure 6 The Training Curve of A3 Experiment

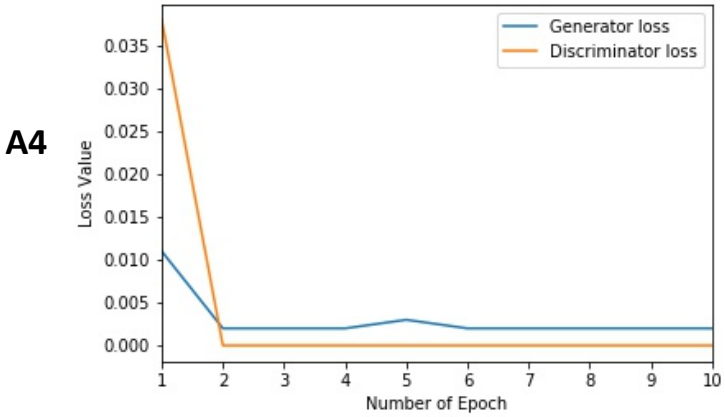

Figure 7 The Training Curve of A4 Experiment

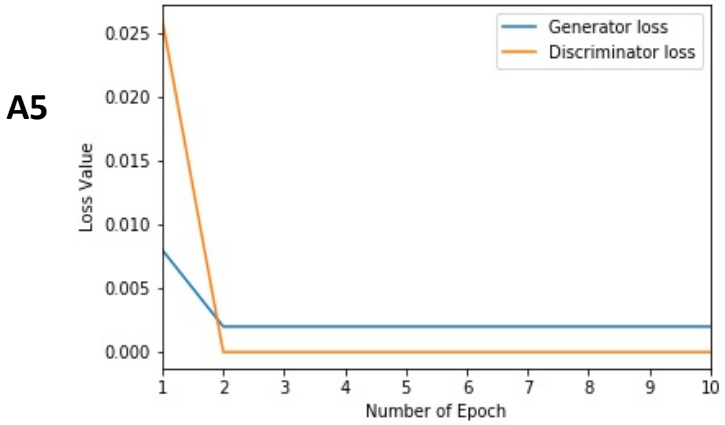

Figure 8 The Training Curve of A5 Experiment

The number of dataset affect loss value of generator and discriminator, as seen from above training curve. Figure 4 shown that loss value of generator and discriminator are relatively bigger when the training process uses a small number of dataset. however loss value are relatively smaller when the training process uses a bigger number of dataset, shown in Figure 8.

Generator produce a relatively real image during training process, so that discriminator can not distinguish between generated image and real image. It means that d-loss will increase, but discriminator will optimize and reduce d-loss in the next training process. Discriminator and generator will reach a balance when generator produce pseudo real images and discriminator can not distinguish them.

4. CONCLUSIONS

This study uses a Conditional Boundary Equilibrium Generative Adversarial Network to increase image resolution and produce super-resolution images. The used size of the input image is 4x smaller than previous studies, with the same size as the output image as the previous study. Although the output image that is produced in this study is not too high resolution, the resulting image has been able to represent the original image. Also, during the training process, the generator and discriminator networks have been able to show good and stable performance. The evaluation process uses a validation dataset and can produce an SSIM value of up to 93%. While the SSIM value generated in the evaluation process using dataset testing reaches a value of 90%. The obtained SSIM value in this study increased by 14% from the previous study.

ACKNOWLEDGMENTS

This study has done in Laboratorium Komputer Dasar of The University of Gadjah Mada, using Graphic Card is NVIDIA GeForce GTX 1080.

REFERENCES

[1] E. Denton, A. Szlam, and R. Fergus, "Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks arXiv : 1506 . 05751v1 [cs . CV] 18 Jun 2015," pp. 1–10, 2015.

[2] I. Goodfellow, "NIPS 2016 Tutorial: Generative Adversarial Networks," 2016, doi: 10.1001/jamainternmed.2016.8245.

[3] 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, doi: 10.1109/CVPR.2017.19.

[4] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "SRLSP: A Face Image Super-Resolution Algorithm Using Smooth Regression with Local Structure Prior," *IEEE Trans. Multimed.*, vol. 19, no. 1, pp. 27–40, 2017, doi: 10.1109/TMM.2016.2601020.

[5] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative Adversarial Networks," *arXiv Prepr. arXiv ...*, pp. 1–9, 2014, doi: 10.1017/CBO9781139058452.

[6] S. Osindero, "Conditional Generative Adversarial Nets," pp. 1–7, 2014.

[7] P. Isola and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017, doi: 10.1109/CVPR.2017.632.

[8] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary Equilibrium Generative Adversarial Networks," pp. 1–10, 2017, doi: 1703.10717.

[9] M. M. and Y. L. Junbo Zhao, "ENERGY-BASED GAN," *Neural Networks*, vol. 61, no. 2014, pp. 32–48, 2015, doi: 10.1016/j.neunet.2014.10.001.

[10] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," pp. 1–10, 2016, doi: arXiv:1504.01391.

[11] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, [Online]. Available: http://arxiv.org/abs/1701.07875.

[12] B. Huang, W. Chen, X. Wu, and C. Lin, "High-quality face image generated with conditional boundary equilibrium generative adversarial networks," *Pattern Recognit. Lett.*, vol. 111, pp. 72–79, 2018, doi: 10.1016/j.patrec.2018.04.028.
[13] W. Shi *et al.*, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," pp. 1–10, 2016.
[14] A. Radford, L. Metz, and S. Chintala, "UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL," pp. 1–16, 2016.