Everyone is A Forensic Artist: Sketch-to-Photo Transformation for Human Face

Daivalentineno Janitra Salim
Department of Information Management
National Taiwan University of Science and Technology
Taipei, Taiwan
daivalentinenojs@gmail.com

Bor-Shen Lin
Department of Information Management
National Taiwan University of Science and Technology
Taipei, Taiwan
bslin@cs.ntust.edu.tw

2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII) | 978-1-6654-2307-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICKII51822.2021.9574719

***Abstract*— There are a lot of crimes happening all over the world every day. Among the criminal acts, homicide is the type of crime with a large number of victims. Occassionally there are witnesses who see the incident and remember the face of the criminal. Thus the police ask them to sketch to find out the suspect. Since the human face is the most significant and informative part of the human body, the sketch of the face is used to identify the suspect with high certainty. However, the suspects may change their facial features by makeup, such as putting on glasses or dyeing hair. If a sketch is converted into photographic images with modified facial features flexibly, the investigation of crime might accelerate effectively.**

Recent research has shown the techniques that transform a sketch of a human face into a photographic image or change the style of a human face according to the designated facial features. However, there has not yet been an integrated architecture to transform pencil sketches to photographic images directly with desired facial features. In addition, there is no fast and automatic evaluation approach that consider multiple metrics of images jointly. It is hence difficult to optimize and select the setting among a few alternatives efficiently. With the limitation, we propose the Forensic GAN, a network of performing the sketch-to-photo transformation and image manipulation according to the designated attributes. A voting mechanism with multiple metrics, including PSNR, SSIM, SCC, and ERGAS, for fast evaluating the image quality jointly was proposed. Accordingly, the training settings, such as the loss function and the number of epochs, can be optimized and selected based on the quality of the synthesized images. The performance of the Forensic GAN was tested, and its potential for the real forensic task has been verified with synthesized pencil images and real hand-drawing sketches.

Keywords—Forensic GAN, Cycle GAN, Star GAN, Data Augmentation, Automatic Evaluation of Image Synthesis

I. Introduction

Many crimes are occurring in various countries every day. The crimes are committed in homicide, robbery, rape, oppression, terrorist attack, and so on. Among the criminal acts, homicide causes a large number of victims according to the Global Burden of Disease, a famous worldwide study on the causes of death and disease distributed in the medical journal [1]. To assist the criminal investigation, announcing the image of the suspect is an effective approach, especially when there are few clues available. The human face is the most informative part of the human body that conveys information such as identity, gender, age, race, and temperament to recognize the identity of a person reliably. Images of human faces are therefore used widely by law enforcement agencies to identify and arrest criminals [2]. Since most people have the capability of remembering human faces, it is usually not difficult to generate a composite sketch of the suspect when there is no witness on the crime scene.

Recent research such as Cycle GAN has shown the capability of transforming the sketched image of a human face into a photographic image or transforming the facial photo into a pencil sketch [3]. Star GAN, on the other hand, was proposed to transform images of the human face into those with specific styles or facial features, such as pointed nose, chubby, and hair color [5]. Such techniques can be combined to fulfill the forensic goal. To train the networks for image transformation, there are a few difficulties. First, real pencil sketches are expensive to obtain as they need to be drawn one by one. For example, there are only 188 real pencil sketches available in the CUHK Face Sketch dataset. Such an amount is usually insufficient for training a good sketch-to-photo transformation network with good quality. It is hence worth investigating the way to augment the pencil-sketch-like images so as to well train the sketch-to-photo transformation network.

In addition, for the image synthesis task, there have not yet been approaches of automatic evaluation of image quality for a lot of synthetic images based on multiple metrics. This makes it difficult to optimize the training. For example, there are a few options of training settings, such as the loss function, the training strategy, the number of epochs, and so on. If there is no automatic approach for fast evaluation of synthetic images produced by different settings, it is hard for the system developer to select a better setting by looking at a large number of images. It becomes even more difficult with multiple metrics of image quality.

Therefore, we proposed the Forensic GAN that combines the Cycle GAN and the Star GAN for the forensic goal. Cycle GAN converts pencil sketches into photographic images, and Star GAN further performs image manipulation according to some designated attributes. The technical issues are summarized below.

(1) As the number of real pencil sketches is limited, a data augmentation approach is proposed to generate a set of sketches of which the quality was verified and filtered by a pencil sketch detector.

(2) A voting mechanism with multiple metrics, including PSNR, SSIM, SCC, and ERGAS, are proposed to evaluate automatically and systematically for a large amount of generated photographic images.

(3) The training settings, such as the loss function and the number of epochs, are tuned and selected according to the quality of a large number of generated images.

(4) The architecture of Forensic GAN, which combines the generator of Cycle GAN and the domain transformer of Star GAN, is proposed to perform the transformation of the human face from pencil sketches to photographic images with designated attributes.

ISBN: 978-1-6654-2307-6

(5) The proposed approaches are verified on the synthesized pencil sketch images in CelebA dataset and real hand-drawing sketches. These techniques have the potential for forensic tasks.

II. Related Works

A. Generative Adversarial Networks

Generative Adversarial Network contains the generator and the discriminator, which are typically denoted as G and D, respectively. The generator produces a fake image X from a noise z. The discriminator is a binary classifier that distinguishes fake or real images. The key success of GAN is the use of adversarial loss to distinguish between real and fake images. This loss function forces the generator to produce the output image that resembles the real image through learning.

B. Cycle GAN

Cycle GAN translates the images across two domains. One is the source domain, and the other is the target domain. The training images in the two domains are not required to be paired [4]. Cycle GAN introduces the cycle loss for reserving the spatial consistency and optimizes jointly four networks, including two domain transformers and two discriminators, based on four objective functions, including two cycle and adversarial losses, respectively.

C. Star GAN

Star GAN performs facial image manipulation and facial expression synthesis. It trains image-to-image translations efficiently from multiple datasets among different domains in a single network [5]. It uses three loss functions, including adversarial loss, domain classification loss, and reconstruction loss.

D. Wasserstein GAN and Deep Regret Analytics

GAN is an effective network model whose concept is from the min-max algorithm in game theory. However, the training of GAN is time-consuming and unstable. Mode collapse arises once in a while due to the sharp gradients of the discriminator function around real data points. One way to handle this problem is to limit the gradient of the discriminator. Wasserstein GAN with gradient penalty and deep regret analytics (DRA) are proposed for that purpose. Wasserstein GAN with gradient penalty makes GAN training stable [6], and deep regret analytics mitigates the mode collapse issue [7].

III. Forensic Generative Adversarial Network

The purpose of Forensic GAN is to transform an image of pencil sketches of the human face into a photographic image with designated attributes. The architecture of Forensic GAN contains three stages, as shown in Fig. 1. Cycle GAN as the first stage is for learning the transformation from the image of pencil sketches into the photographic image, as shown in Fig. 1(a). Star GAN as the second stage is for learning the image manipulation that transforms the facial images with assigned attributes, as shown in Fig. 1(b). Forensic GAN as the third stage finally combines the generator of Cycle GAN and the generator of Star GAN so as to transform the pencil sketches into the photographic image with designated attributes serially in one shot.

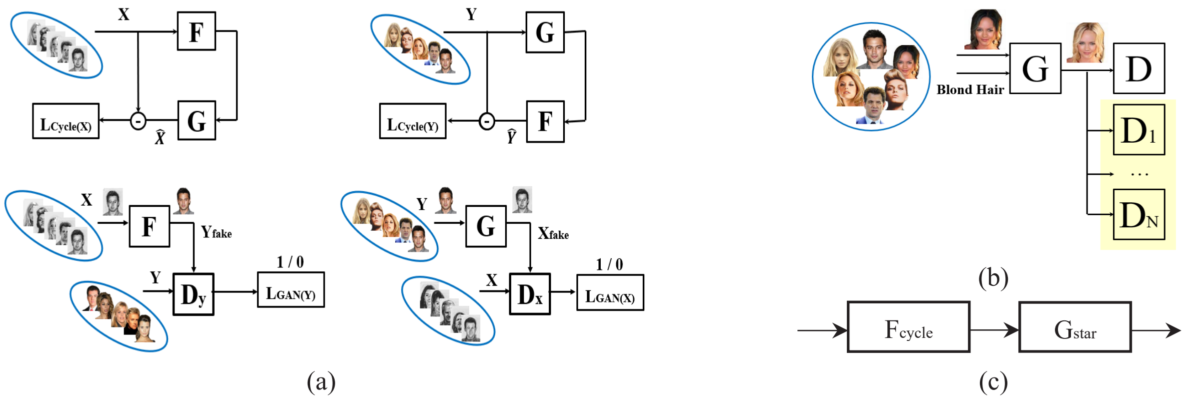

Fig. 1. Forensic GAN architecture of three stages. (a) Cycle GAN for sketch-to-photo transformation (b) Star GAN for image manipulation (c) Forensic GAN combines the generators of Cycle GAN and Star GAN to transform the pencil-sketch-like image into photographic image with designated attributes.

A. Data Augmentation

To train sketch-to-photo transformation based on Cycle GAN, it is required to provide sufficient training data in both domains. It is simple to collect a large number of facial photos, but it is not easy to obtain a lot of real sketches with good quality. In this study, the CelebA dataset and CUFSF dataset are selected as the photo and pencil sketch datasets, respectively. Since the number of real sketches in the CUFSF dataset is limited, not all of the generated pencil sketch images have good quality. To improve the quality of the images, the non-face background was first removed from the photographic image to avoid interference by the background. In addition, a binary pencil sketch detector is trained to tell whether a gray-level image looks like a pencil-like image or not. Every pencil sketch is evaluated with the pencil sketch detector, and the average percentage for all generated images is 79.75%. The top 2,500 synthesized pencil sketch images with the highest evaluation scores are selected and included in the sketch dataset for learning the sketch-to-photo transformer of Cycle GAN.

B. Loss Function

Adversarial Loss To match the distribution of the generated images to the distribution of the data in the target domain, the adversarial loss is included in the loss function as below.

$$L_{adv} = E_{y\sim p_{data}(y)}[(D_Y(y)-1)^2] + E_{x\sim p_{data}(x)}[(D_Y(G(x)))^2] \quad (1)$$

where $G(x)$ is the output of the generator which converts the image of domain X into the image of domain Y. $G(x)$ is a fake image and should be rejected by the discriminator D_Y. Therefore, $D_Y(G(x))$ should approach 0 as possible. $D_Y(y)$ should approach 1 because y is an image of domain Y.

Cycle Consistency Loss Cycle consistency loss is to reserve the spatial consistency for the images when applying both the transformations, *G* and *F*. For every image *y* of domain *Y*, when it is translated into domain X and converted back to the original domain Y, the reconstructed image should be spatially correlated to the original image. That is, G(F(y)) ≈ y. Similarly, the domain X holds that F(G(x)) ≈ x. Therefore, the objective function is written with the L1-norm as below.

$$L_{cyc} = E_{x\sim p_{data}(x)}\left[\|F(G(x))-x\|_1\right] + E_{y\sim p_{data}(y)}\left[\|G(F(y))-y\|_1\right] \quad (2)$$

Domain Classification Loss The goal of Star GAN is to transform an input image with designated attributes. For each attribute, a corresponding domain classifier is trained to discriminate the domain based on the classification loss.

$$L_{cls} = E_{x,c}\,[-\log D_{cls}(c|x)] \quad (3)$$

where $D_{cls}(c|x)$ is a probability over domain labels that are output from the discriminator. The domain classifier serves as an extra discriminator that trains the generator how to convert the input image with the designated attribute *c*. During training, it is possible to convert multiple attributes, and many domain classifiers are enabled to train the generator.

Reconstruction Loss To make the generated image preserve the spatial content of the input image during transformation in Star GAN, the reconstruction loss is introduced as below.

$$L_{rec} = E_{x,c,c'}\,[\|x - G(G(x,c),c')\|_1] \quad (4)$$

where c is the label of the target attribute and is the label of the original attribute. If the generator outputs the generated image $G(x,c)$ and converts it back to the original domain, the reconstructed image should be close to the original image x.

Deep Regret Analytics Loss Based on this idea of deep regret analytics, regret minimization is used to avoid the mode collapse and implemented as below.

$$L_{GP} = \lambda\,.\,E_{x\sim P_{real},\delta\sim N_d(0,c)}[\|\nabla_x D_\theta(x+\delta)\|-1]^2 \quad (5)$$

where λ = 10 and c = 10. The regret minimization prioritizes the regret such that it does not grow in proportion to time. DRAGAN hypothesizes that the sudden drop of the inception score is related to the gradient's norm, so the gradient penalty is added to the cost function. It provides asymptotic convergence of GAN training in the non-parametric limit and optimizes the discriminator at each step.

C. Dataset and Settings

CelebA Color Images. The CelebA dataset contains 202,599 facial photos of celebrities, in which each image has 40 binary attributes. We randomly select 2,000 images as a testing set and 2,500 images as a training set.

CelebA Pencil Sketch Images. Since the number of real pencil sketches is not sufficient for training the sketch-to-photo transformer of Cycle GAN, the data augmentation approach is utilized to produce more synthesized pencil-like images. 202,599 photos in the CelebA dataset were converted into such images using a pre-trained Cycle GAN. However, not all the output images have good quality, and a few of them look simply similar to the gray-scaled images rather than pencil sketches. The pencil sketch detector is therefore trained to evaluate those synthetic images. Top 2,000 images among all the synthetic images are selected to augment the sketch dataset according to the evaluation scores from the pencil sketch detector.

Parameters. All models are trained using Adam [8] with $\beta 1$ = 0.5 and $\beta 2$ = 0.999, and the batch size is set to 16. During training, the generator is updated once when the discriminator is updated five times [9]. The learning rate is 0.0001 for the first 10 epochs and decays linearly to 0 over the next 10 epochs.

D. Voting Mechanism with Multiple Metrics

During testing, every pencil sketch is used as input to produce the photographic image with the generator of Cycle GAN. However, there are training settings such as loss function and the number of epochs, so a list of output images is generated for each test image. Since there are usually a lot of testing images, it is hard to know which setting is better without automatic evaluation. Moreover, there are a few metrics frequently used for the evaluation of image quality, and how to integrate multiple metrics efficiently has not yet been solved. In this study, a voting mechanism with multiple metrics is proposed to deal with the aformentioned issues. Four metrics, including PSNR, SSIM, ERGAS, and SCC, are used individually to rank the generated images for each test image, and each metric awards the top three images with 3, 2, and 1 votes, respectively. When the voting is conducted on all test images, the number of votes for each setting accumulates, and the best settings are finally determined according to the total number of votes. In this way, automatic evaluation of image quality with multiple metrics for a large number of synthetic images is performed efficiently to optimize the training settings.

IV. EXPERIMENTS

A. Optimization of Training Settings

First, two objective functions, loss-sensitive function and vanilla loss of Cycle GAN are compared, and the number of epochs, 50, 100, 150, and 200, are tested, respectively. Therefore, for every pencil sketch, eight photographic images are generated. The voting mechanism is conducted for the eight images corresponding to the eight settings by 2 objectives and 4 numbers of epochs. Four metrics, PSNR, SSIM, ERGAS, and SCC, serve as the judges that rank the eight candidates and award their votes to the top-three images among eight. After the voting is conducted for all the testing images, the votes from every metric are distributed to the eight settings, and it is then ready to tell which settings are superior by counting the accumulated votes. Table 1 shows the accumulated votes for the settings from each metric and all metrics. Table 1 shows that most metrics of image quality favor the loss-sensitive function more than the vanilla loss. In addition, the best number of epochs appears to be 200 which obtains 8,458 votes from all metrics. The proposed voting mechanism is an efficient approach for evaluating a large number of outputs automatically. When there are a few alternatives to training settings, it is not only inefficient but impossible to select a good setting reliably by simply looking at a lot of synthetic images. With the proposed approach, the training setting is selected objectively and efficiently based on the image quality of the images with multiple desired metrics. For example, the loss-sensitive function and training by 200 epochs seem to be good options since the setttings generate more images with higher quality (Table 1). Therefore, such settings is used for the Forensic GAN later.

TABLE 1. AUTOMATIC EVALUATION OF IMAGE QUALITY FOR CYCLE GAN BASED ON THE VOTING MECHANISM WITH MULTIPLE METRICS. THE TRAINING SETTINGS INCLUDE THE NUMBER OF EPOCHES, 50 THROUGH 200, AND THE OBJECTIVE FUNCTIONS, LOSS SENSITIVE OR VANILLA.

	Loss Sensitive				Vanilla			
	50	100	150	200	50	100	150	200
PSNR↑	927	422	583	1855	220	696	985	312
SSIM↑	851	601	968	2471	137	318	383	272
ERGAS↓	985	577	710	1627	200	672	943	286
SCC↑	203	1590	1494	2505	19	91	13	85
Total	2966	3190	3755	8458	576	1777	2324	955

B. *Classification Accuracy of Attributes*

In Star GAN for image manipulation, we compare the training strategies of Wasserstein gradient penalty, denoted as WGP, and Wasserstein gradient penalty plus deep regret analytics, denoted as WGP+DRA, respectively. 2,500 real color images of facial photos are selected for training and 1,000 images for testing randomly from the CelebA dataset. The accuracy of attribute classification by the domain classifier is used as the performance indicator of image manipulation. The testing is conducted on Star GAN trained with WGP and WGP+DRA, respectively, and the accuracies of attribute classification are shown in the middle two columns Table 2. The first column of Table 2 displays the classification accuracies for real images for reference. Star GAN with WGP+DRA achieves higher accuracies than Star GAN with WGP for all attributes, and the performance gap between is 2.93% on average, which means regret minimization is effective for improving the Wasserstein gradient penalty. Since WGP+DRA is superior to WGP in all cases, the generator of Star GAN trained with WGP+DRA is finally selected to build the Forensic GAN.

In addition, the Forensic GAN needs to transform the pencil sketches into photographic images and perform image manipulation. The classification accuracies of all the attributes for the synthesized images of Forensic GAN are displayed in the last column of Table 2. Though the Forensic GAN uses the same generator as Star GAN, the image manipulation of the Forensic GAN takes the photographic image transformed from pencil sketches as input, instead of the real facial image. The Foresic GAN achieves the classification accuracy of 65.53% on average, which is lower than Star GAN that uses real photos as input (Table 2). For such attributes as Back Hair and Eyeglasses, the Forensic GAN outperforms the Star GAN. In summary, with the data augmentation of pencil sketches and the optimization of training settings, Forensic GAN based on WGP+DRA performs sketch-to-photo transformation and image manipulation with the performance compatible with Star GAN.

C. *Forensic Application*

To investigate the real effects of the Forensic GAN, it is tested on 30 photos extracted from the CelebA dataset, and 30 criminal photos collected from Google. The testing results for the photos of the CelebA dataset and criminal dataset are displayed in Figs. 2 and 3, respectively. The results show that the Forensic GAN transforms sketches into photographic images and performs image manipulation successfully. The potential has been verified for the forensic task on the real hand-drawing sketches.

TABLE 2. CLASSIFICATION ACCURACY OF IMAGE MANIPULATION FOR STAR GAN AND FORENSIC GAN (WITH WGP+DRA).

Attributes	**Real Image (%)**	**Star GAN +WGP (%)**	**Star GAN +WGP +DRA (%)**	**Forensic GAN (%)**
Attractive	90.21	64.21	66.06	65.18
Bald	90.24	67.25	72.14	68.35
Bangs	85.72	65.99	67.10	65.22
Black Hair	86.86	67.16	72.52	73.22
Blond Hair	95.87	85.79	85.89	85.25
Brown Hair	90.37	71.25	74.63	71.94
Chubby	87.06	71.12	74.01	72.45
Double Chin	86.93	70.25	73.32	72.24
Eyeglasses	89.1	71	78.70	80.82
Mustache	87.67	67	69.27	66.29
Pale Skin	87.03	68.17	75.02	75.21
Sideburns	82.66	56.91	59.17	58.31
Smiling	82.96	57.27	59.23	56.17
Wavy Hair	86.45	67.26	69.88	65.23
Wearing Hat	89.67	65.58	70.39	65.04
Average	**86.27**	**64.25**	**67.18**	**65.53**

V. CONCLUSIONS

In this paper, we proposed the Forensic GAN, an image-to-image translation model that transforms the pencil sketch image of the human face into photographic images based on the designated attributes. Data augmentation provides sufficient sketch data. A voting mechanism with multiple metrics evaluates a large number of synthetic images, which optimizes the training settings to obtain the output images with good quality. The potential of Forensic GAN for the real forensic task is verified on the CelebA data and the real sketches.

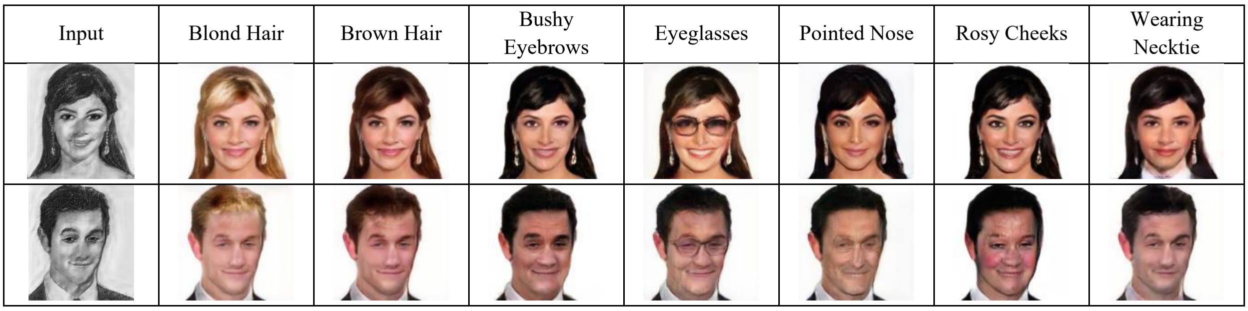

Fig. 2. Sketch-to-photo transformation results trained with the CelebA dataset via Forensic GAN. The input is a pencil sketch image and Forensic GAN can transform the input image based on the designated attributes.

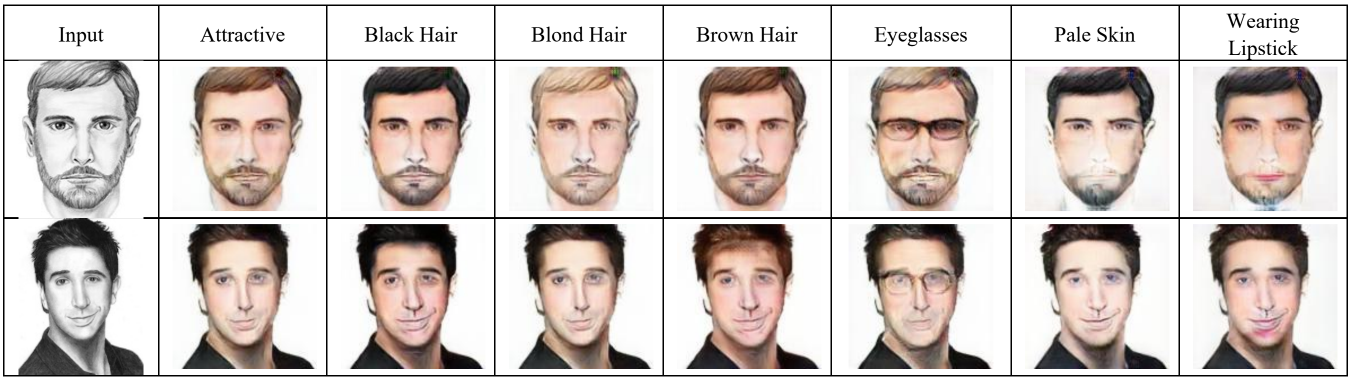

Fig. 3. Sketch-to-photo transformation results on the real sketches via Forensic GAN.

REFERENCES

[1] GBD 2017 Causes of Death Collaborators (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1736-1788. https://doi.org/10.1016/S0140-6736(18)32203-7

[2] Saraiva, Renan & Castilho, Goiara & Nogueira, Raiane & Coelho, Letícia & Alarcão, Luciana & Lage, Jade. (2017). Eyewitnesses memory for faces in actual criminal cases: An archival analysis of positive facial composites. Estudos de Psicologia (Natal). 22. 247-256. 10.22491/1678-4669.20170025.

[3] Yu, J., Xu, X., Gao, F., Shi, S., Wang, M., Tao, D., & Huang, Q. (2020). Toward Realistic Face Photo-Sketch Synthesis via Composition-Aided GANs. *IEEE Transactions on Cybernetics*.

[4] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).

[5] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8789-8797).

[6] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*.

[7] Kodali, N., Abernethy, J., Hays, J., & Kira, Z. (2017). On convergence and stability of gans. *arXiv preprint arXiv:1705.07215*.

[8] D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.