# DrawGAN: Multi-view Generative Model Inspired By The Artist's Drawing Method *

Bailin Yang[1(✉)][0000-0003-1754-5595], Zheng Chen[1][0009-0002-3164-0492], Frederick W. B. li[2][0000-0002-4283-4228], Haoqiang Sun[1][0009-0000-6135-8936], and Jianlu Cai[1][0009-0004-7290-5308]

[1] Zhejiang GongShang University, China
ybl@zjgsu.edu.cn, {1476464203,875968520,963451397}@qq.com
[2] University of Durham, United Kingdom
frederick.li@durham.ac.uk

**Abstract.** We present a novel approach for modeling artists' drawing processes using an architecture that combines an unconditional generative adversarial network (GAN) with a multi-view generator and multi-discriminator. Our method excels in synthesizing various types of picture drawing, including line drawing, shading, and color drawing, achieving high quality and robustness. Notably, our approach surpasses the existing state-of-the-art unconditional GANs. The key novelty of our approach lies in its architecture design, which closely resembles the typical sequence of an artist's drawing process, leading to significantly enhanced image quality. Through experimental results on few-shot datasets, we demonstrate the potential of leveraging a multi-view generative model to enhance feature knowledge and modulate image generation processes. Our proposed method holds great promise for advancing AI in the visual arts field and opens up new avenues for research and creative practices.

**Keywords:** Unconditional GANs · AI art · few-shot dataset · multi-view generative model

## 1 Introduction

GANs have revolutionized machine learning by generating novel visual content through modeling high-dimensional distributions They have excelled in various applications such as image translation, frame prediction, 3D modeling, and sketch-to-image synthesis [20,14,26,15]. However, training GANs is challenging due to the non-convex game nature and high-dimensional parameter space, leading to mode collapse and training instability [1,32].

While unconditional GANs perform well on class-specific datasets, they struggle with generating outline, gray, and color views simultaneously [19,31], which

* Supported by National Natural Science Foundation of China NO 62172366
Supported by Key Laboratory of Electronic Business and Logistics Information Technology of Zhejiang Province KF202201

limits the generality of image generation. State-of-the-art methods like StyleGAN and StyleGAN2 offer reliable performance but require significant computational resources [12,13]. FastGAN addressed this by proposing a skip-layer excitation module and a self-supervised discriminator [18].

However, current GANs have limitations in evaluating a single view of an image, leading to mode collapse and neglect of important image features. To address this, we propose DrawGAN, an unconditional generative model inspired by the artist's drawing process. In the artist's process, coloring is applied on top of object shapes, indicating that line drawing and shading contain valuable knowledge for coloring. DrawGAN aims to enhance the detection of diverse descriptors and capture more image features by utilizing a multi-component approach.

Our approach models the image generation process using outline, gray, and color components, with outline and gray integrated into the color component. We believe these components are mutually dependent and complementary. The DrawGAN model incorporates a novel multi-view generator capable of generating images at three levels: outline, gray, and color. This approach aims to produce high-quality synthesized images with improved diversity and fidelity. Our main contributions are as follows:

1. Proposal of the integration of the artist's painting concepts into the unconditional generative adversarial network, resulting in the novel DrawGAN model structure.
2. Development of DrawGAN, which incorporates a three-view generator and a multi-discriminator. This architectural design empowers the model to effectively capture diverse image information, including multiple view features, and mitigate the mode collapse issue.
3. Execution of experiments on few-shot datasets to validate and demonstrate the robustness of DrawGAN in generating good-quality images.

## 2 RELATED WORK

### 2.1 Image Generation and Synthesis

In recent years, the field of image generation and synthesis has witnessed numerous variations of Generative Adversarial Networks (GANs) aiming to improve the quality of generated samples.

**Few-Shot Image Generation:** Generating high-quality images with GANs in a few-shot scenario is challenging due to the need of large datasets during training (typically 50,000 to 100,000 images). When training data is limited, the discriminator can easily overfit, leading to inadequate feedback for the generator.

To address this challenge, DiffAug and StyleGAN-Ada rely on data augmentation techniques to prevent discriminator overfitting [34,11]. ContraD introduced a strong data augmentation strategy by learning a contrastive representation compatible with GANs [9]. Another recent approach utilizes LeCam divergence, minimizing an $f$-divergence through modulation of discriminator predictions [28]. However, while these methods have shown success in capturing information from single views, they do not support the generation of multiple image versions (e.g., outline and gray) within the same GAN architecture.

**Image-to-Image Translation:** Image-to-image translation involves converting an input image into a different output form. Establishing the correlation between input and output images is a significant challenge. When the network fails to capture this correlation, the generator may cheat by ignoring the input, resulting in output images similar to the truth images. StarGAN and its successor, StarGAN v2, propose a single network model for image-to-image translations across multiple domains, requiring paired image datasets from these domains [3,4]. Nazeri et al. divide image inpainting into two steps: predicting missing region edges and completing the image using the predicted edges [22]. While DrawGAN shares similarities with their idea, we differ by generating three views (outline, gray, and color) using a single generator, whereas they employ two separate generators.

Differing from the aforementioned works, our focus lies in unconditional image generation, which removes the constraint of paired images as inputs. Notably, our work introduces a novel multi-view generative network capable of simultaneously generating outline, gray, and color versions of images. This advancement enables the synthesis of diverse and comprehensive images.

### 2.2 AI Art

AI technology advancement for digital art was marked by the milestone of image style transfer [6]. This technique utilized CNNs to generate stylized images by separating and recombining the "content" and "style" of an image.

Building upon GAN for generating creative images, Elgammal et al. introduced Creative Adversarial Networks (CAN) [5]. CAN incorporates two feedback signals from the discriminator to the generator: classification of "art or not art" and the ability to classify the generated art into established styles. By optimizing these signals, CAN enables the generation of creative images that deviate from established styles while conforming to the distribution of art.

Yi et al. proposed APDrawingGAN, a GAN-based approach for generating artistic portrait drawings from face images [30]. This method employs hierarchical generators and discriminators, leveraging both global and local convolutional networks to extract facial features. These features are then fused to produce the final output. In contrast, Li et al. presented a two-stage method for colorizing line drawings using GANs [16]. The initial stage generates a color draft that aggressively applies colors to enrich variety, despite potential errors such as color mistakes, bleeding, and blurring/distortion. The second refinement stage utilizes a synthetic paired image dataset to address these errors.

The novelty of our DrawGAN lies in its ability to generate three distinct representations of an image (outline, grayscale, and color) using a single multi-view generative network. Unlike existing methods that produce only one representation, DrawGAN expands the possibilities of image synthesis and offers the exploration of diverse visual representations. This advancement in AI art facilitates new creative possibilities and opens avenues for artistic expression.

## 3 Method

Inspired by the art drawing process, we propose DrawGAN, an architecture capable of generating outline, gray, and color images (Figure 1). Training our network requires datasets with all three views, but existing datasets typically provide only the color view. To overcome this limitation, we apply traditional algorithms to extract the outline and gray views from color images, creating our own dataset. We use the canny algorithm for edge extraction to generate the outline view and the Floyd-Steinberg dithering algorithm to approximate brightness and obtain the gray view.

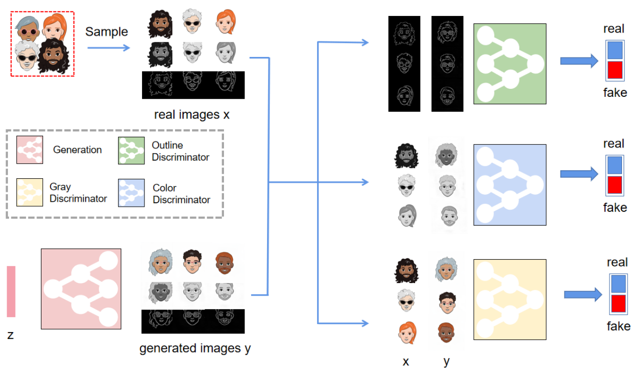



**Fig. 1.** DrawGAN utilizes a multi-view generator with three outputs, feeding to three corresponding discriminators, which are trained with distinct loss functions for discriminating real and fake outputs of the outline, gray, and color views of images, respectively.

### 3.1 Loss Formation

In a vanilla GAN, the discriminator D and the generator G engage in a minimax game to guide the generator in synthesizing noise vectors into realistic images. For the sake of simplicity, the equation can be expressed as:

$$\mathcal{L}_D = \underset{x\sim P_d}{E}[\log D(x)] + E_{z\sim p_z}[\log(1 - D(G(z)))] \tag{1}$$

$$\mathcal{L}_G = \underset{z\sim N}{E}[\log D(G(z))] \tag{2}$$

There have been numerous studies on GAN loss functions since the introduction of GAN, including wgan [2], wgan-gp [7], and hinge loss [17]. According to various experiments, hinge loss is generally considered the most stable and efficient approach for unconditional image generation [17], [27]. Therefore, we use the hinge version of the adversarial loss to train our D and G iteratively.

$$\mathcal{L}_D = -Ex \sim P_d[\min(0, -1 + D(x))] - E_{z\sim N}[\min(0, -1 - D(G(z)))] \tag{3}$$

$$\mathcal{L}_G = -Ez \sim N[D(G(z))] \tag{4}$$

Our proposed architecture, in contrast to existing GAN architectures, includes a multi-view generator that produces three distinct outputs. This model architecture not only follows the artist drawing process but also adheres to the GAN training mechanism. As shown in Fig. 1, our model comprises a generator with three outputs that are linked to three distinct discriminators.

The generator in our proposed DrawGAN architecture is different from traditional GAN generators, which usually generate a single-view image $y$. Instead, our generator produces an image set $y_1, y_2, y_3$ for the outline, gray, and color views, respectively, from a noise vector $z \sim N$, where $y_1, y_2, y_3 = G(z)$. Similarly, for training our network, we use image sets of outline, gray, and color views $x_1, x_2, x_3$ from real images. Corresponding discriminators $D_1, D_2$, and $D_3$ are constructed to distinguish real and fake views, respectively. Hence, Equation 3 can be rewritten as:

$$\mathcal{L}_{D_i} = -E[\min(0, -1 + D_i(x_i))] - E[\min(0, -1 - D_i(y_i))] \tag{5}$$

FastGAN [18] proposed reconstructive training to ensure that D can extract more comprehensive representations from inputs, covering both overall composition and detailed textures, with minimal additional computational cost. Our discriminator has adopted this reconstructive training technique, such that:

$$\mathcal{L}_{recon} = E[||g(f) - T(x)||] \tag{6}$$

Therefore, equation 5 is further rewritten as:

$$\begin{gathered} \mathcal{L}_{D_i} = -E[\min(0, -1 + D_i(x_i))] \\ -E[\min(0, -1 - D_i(y_i))] + \mathcal{L}_{recon} \end{gathered} \tag{7}$$

The final losses for generators and discriminators are formulated as follows:

$$\mathcal{L}_D = \mathcal{L}_{D_1} + \mathcal{L}_{D_2} + \mathcal{L}_{D_3} \tag{8}$$

$$\mathcal{L}_G = -E[D(y_1)] - E[D(y_2)] - E[D(y_3)] \tag{9}$$

### 3.2 Model Architecture

Based on the equations provided, we propose a multi-view generator architecture with three outputs, as illustrated in Fig. 2. The input noise vector $z$ is first transformed into a 4x4 feature space through a Transpose Convolution (4x4) and a Convolution (3x3). We then employ a process of upsampling followed by two 3x3 convolutions to improve the feature space resolution, which is commonly used in most GAN architectures [23,18,10]. Finally, we define a mapping function $r$ to generate the output images corresponding to the outline, gray, and color views. The mapping function $r$ is implemented using a convolution (1x1) layer to project the feature space onto the image space. To enhance the stability of the training, we apply spectral normalization [21] to all of the convolutional layers.

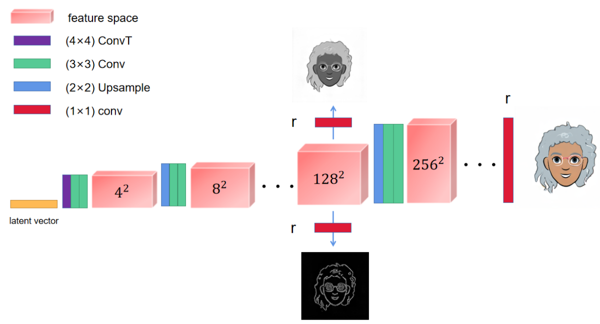



**Fig. 2.** The generator architecture uses nearest-neighbor interpolation and a 3x3 convolution for upsampling. The feature space of 128 is processed by function r to generate outline and grayscale views, while the final feature space is for color view generation.

Our generator is designed to mimic the artist's drawing process by first generating the line drawing (outline) view and shading (gray) view, and then generating the color view, as shown in Fig. 2. We use a 128x128 resolution feature space to output the outline and gray views and the final feature space to output the color view of the image. The design of our model follows the forward propagation mechanism of neural networks.

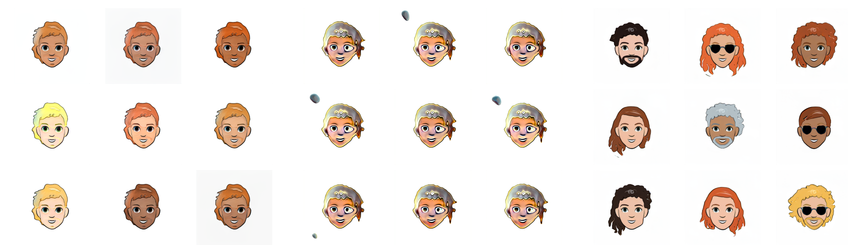
As shown in Fig. 1, our DrawGAN multi-discriminator comprises three modules, namely outline discriminator, gray discriminator, and color discriminator. They provide useful feedback to the generator by comparing the generated image set $y_i$ with the real image set $x_i$. We adopt the FastGAN model [18] to construct the architecture of our multi-discriminator, where the color discriminator remains unchanged, and the number of channels for outline and gray discriminator is reduced to half that for the color discriminator.

In terms of the discriminator's auto-encoding reconstruction, we adopt different reconstruction loss functions for different views. The color image reconstruction loss uses the perceptual similarity metric [33], while the outline and gray image reconstruction loss use the structural similarity index (SSIM) [29]. The DrawGAN generator $G$ produces three fake samples $Y_i$ from a noise vector $z$, where the number of output channels for outline and gray is 1. $X_i$ and $Y_i$ are then fed into the multi-discriminator for training DrawGAN using the loss functions defined in Equation 8 and Equation 9.

## 4 Experiment

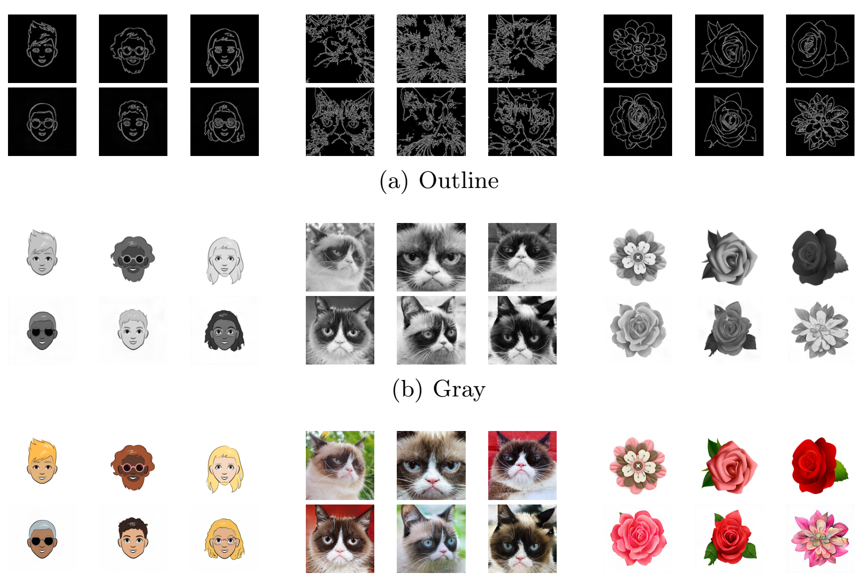
### 4.1 Datasets and Evaluation Metrics

To align with the artist drawing process, our model focuses on extracting clear line drawings (outlines) from the datasets. While line drawings from natural images tend to be chaotic, those from artificial images are typically clearer.



(a) StyleGAN2 (b) FastGAN (c) DrawGAN

**Fig. 3.** The qualitative results of the cartoon face dataset clearly demonstrate that both StyleGAN2 [13] and FastGAN [18] are plagued by mode collapse, while our DrawGAN produces more realistic results. These findings indicate that DrawGAN achieves state-of-the-art performance on this dataset.



(a) Outline

(b) Gray

(c) Color

**Fig. 4.** The comparison between real images (top rows) and the generated images (bottom rows) highlights DrawGAN's ability to generate random and uncurated images with three different views (outline, gray, and color) on high-resolution datasets.

Consequently, our image generation task primarily concentrates on artificial images. However, we also conduct experiments with natural images to showcase the generality of DrawGAN. Overall, our experiments cover multiple datasets spanning diverse content categories.

**Artificial images:** We evaluate our method on public datasets utilized in FastGAN [18]. This includes 1000 paintings from WikiArt (wikiart.org) with a resolution of 1024x1024, 833 Pokemon images with a resolution of 1024x1024 (pokemon.com), 500 cartoon face images from the cartoon face dataset [24] with a resolution of 512x512, and 125 flower images from pngimg (pngimg.com) with a resolution of 512x512.

**Natural images:** DrawGAN is evaluated on datasets, including the widely-used AFHQ Cat dataset [25], consisting of 160 cat face images at 256x256 resolution, the 100-Shot-Panda dataset, containing 100 panda face images at 256x256 resolution, the 100-Shot-Grumpy-cat dataset, comprising 100 grumpy cat faces at 256x256 resolution [34], and a set of 60 shell images at 1024x1024 resolution.

**Evaluation metrics:** We adopt the Fréchet Inception Distance (FID) [8] as our evaluation metric to assess the quality of the generated images. FID measures the Wasserstein distance between the feature space representations of real and generated images, providing an evaluation of both quality and diversity.

**Comparative model:** DrawGAN is compared with StyleGAN2 [13], a powerful but resource-intensive unconditional model, and FastGAN [18], an architecture designed for low-data image synthesis. Official PyTorch implementations of these models were used, and they were trained with the reported best configurations. All experiments were conducted on a single RTX-3060 GPU. While some evaluation results may vary due to equipment differences, we present the best results obtained in our experiments.

### 4.2 Experiments on Artificial Image Datasets

Here, we present the results of our experiments on four artificial image datasets. Table 1 presents a comparison of our model with StyleGAN2 and FastGAN at various resolutions on these datasets. Fig. 3 shows the qualitative results of cartoon faces, where we observe that both StyleGAN2 and FastGAN suffer from mode collapse, while DrawGAN generates more realistic results. In other words, with artificial datasets that have clear outlines, DrawGAN significantly outperforms state-of-the-art methods in terms of image synthesis quality. However, DrawGAN still outperforms state-of-the-art methods, albeit with smaller improvements, when dealing with datasets that lack clear outlines.

### 4.3 Experiments on Natural Image Datasets

We present our experiment results conducted on four natural image datasets. The results between our model and StyleGAN2 and FastGAN at various resolutions on multiple datasets are summarized in Table 2. Our model demonstrates impressive performance in generating images of natural scenes, showcasing its

**Table 1.** Comparing with state-of-the-art models over artificial image datasets trained with 500 (cartoon face), 1000 (painting), 125 (flower), and 833 (pokemon) samples, DrawGAN consistently outperforms them and mitigates mode-collapse on generator.

| Datasets | Cartoon face | Painting | Flower | Pokemon |
|---|---|---|---|---|
| Image Number | 500 | 1000 | 125 | 833 |
| Resolution | 512 | 1024 | 512 | 1024 |
| StyleGAN2 | 65.4 | 74.56 | 112.10 | 60.12 |
| FastGAN | 145.08 | 45.08 | 81.66 | 57.19 |
| Our Method | **27.4** | **42.35** | **75.32** | **46.90** |

versatility. However, as shown in the middle section of Fig. 4, the outline extraction for natural images is not as clear, indicating that the lack of clarity in line drawing might be a contributing factor to the relatively modest improvement achieved by our method. It is important to note that the artist drawing process, which is the focal point of our paper, may not entirely align with the process of creating natural scene images. In reality, natural objects may lack explicit outlines, and the outlines we perceive are often formed by variations in environmental illumination or disparities in colors among foreground and background objects, as well as different parts of an object. Such outlines are implicitly created by gray and color features, which can be ambiguous or untidy in nature.

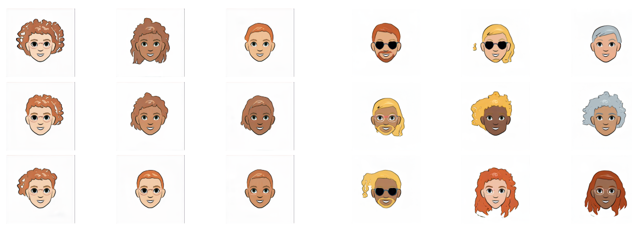
**Table 2.** Comparing with state-of-the-art models over natural image datasets trained with 100 (Grumpy Cat and Panda), 160 (AFHQ Cat), and 60 (shell) samples, DrawGAN's performance slightly outperforms the state-of-the-art.

| Datasets | AFHQ Cat | Shell | Panda | Grumpy cat |
|---|---|---|---|---|
| Image Number | 160 | 60 | 100 | 100 |
| Resolution | 256 | 1024 | 256 | 256 |
| StyleGAN2 | 42.44 | 220.45 | 12.06 | 27.08 |
| FastGAN | 35.11 | 155.47 | 10.03 | 26.65 |
| Our Method | **34.32** | **114.62** | **9.56** | **24.48** |

### 4.4 Discussion

Unconditional GAN models typically generate images from a single view, which may result in limited contour and shading information. In contrast, our multi-view approach captures diverse information, enhancing the image representation. The outline and gray components play crucial roles, with the outline capturing strong gradient changes and the gray conveying color intensity and shading. Our model effectively incorporates both features, enabling a wider range of image information to be considered. Fig. 4 provides a visual comparison of the generated data types with real data.

We conducted an experiment to explore an alternative approach of extracting outline and gray features directly from the final generated image, after the color view. However, the results in Fig. 5 demonstrate that the original setting of DrawGAN outperforms this alternative approach in generating high-quality results. This finding confirms the rationality of following the artist's drawing process. However, generating the outline and gray views after the color view requires adjusting the color feature space, resulting in a decrease in color richness.



(a) (b)

**Fig. 5.** Right group of 9 color images: Best results obtained by generating outline and gray views after the color view (FID 65.5). Left group of 9 color images: Best results obtained based on original setting of DrawGAN (FID 27.4).

## 5 Conclusion

DrawGAN, a novel multi-view generative model, addresses the challenge of over-fitting in limited data image generation tasks by introducing a multi-view generator and three discriminators. The multi-view design enriches the feature space, allowing the generator to capture complex structures and patterns. The discriminators focus on different views, providing diverse information to mitigate over-fitting. The sequential generation of outline, gray, and color views enhances model stability and gradient flow. Experimental results on multiple few-shot datasets demonstrate that DrawGAN outperforms state-of-the-art methods, showcasing its effectiveness in image generation. However, limitations are observed in generating backgrounds with diverse colors, due to limitations in outline and gray views. The equal weighting of views in the loss function dilutes background color information. DrawGAN introduces a novel approach to GANs, with potential for driving advancements in the field of digital art.

## References

1. Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.

2. Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
3. Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.
4. Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8188–8197, 2020.
5. Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*, 2017.
6. Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
7. Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.
8. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.
9. Jongheon Jeong and Jinwoo Shin. Training gans with stronger augmentations via contrastive discriminator. *arXiv preprint arXiv:2103.09742*, 2021.
10. Animesh Karnewar and Oliver Wang. Msg-gan: Multi-scale gradients for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7799–7808, 2020.
11. Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, 33:12104–12114, 2020.
12. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
13. Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.
14. Yong-Hoon Kwon and Min-Gyu Park. Predicting future frames using retrospective cycle gan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1811–1820, 2019.
15. Junsoo Lee, Eungyeup Kim, Yunsung Lee, Dongjun Kim, Jaehyuk Chang, and Jaegul Choo. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5801–5810, 2020.
16. Chengze Li. Two-stage sketch colorization. *ACM Transactions on Graphics*, 2018.
17. Jae Hyun Lim and Jong Chul Ye. Geometric gan. *arXiv preprint arXiv:1705.02894*, 2017.
18. Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In *International Conference on Learning Representations*, 2020.

19. Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.
20. Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
21. Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
22. Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv preprint arXiv:1901.00212*, 2019.
23. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
24. Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. In *Domain Adaptation for Visual Understanding*, pages 33–49. Springer, 2020.
25. Zhangzhang Si and Song-Chun Zhu. Learning hybrid image templates (hit) by information projection. *IEEE Transactions on pattern analysis and machine intelligence*, 34(7):1354–1367, 2011.
26. Ayush Tewari, Mohamed Elgharib, Gaurav Bharaj, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. Stylerig: Rigging stylegan for 3d control over portrait images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6142–6151, 2020.
27. Dustin Tran, Rajesh Ranganath, and David M Blei. Deep and hierarchical implicit models. *arXiv preprint arXiv:1702.08896*, 7(3):13, 2017.
28. Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing generative adversarial networks under limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7921–7931, 2021.
29. Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
30. Ran Yi, Yong-Jin Liu, Yu-Kun Lai, and Paul L Rosin. Apdrawinggan: Generating artistic portrait drawings from face photos with hierarchical gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10743–10752, 2019.
31. Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.
32. Dan Zhang and Anna Khoreva. Pa-gan: Improving gan training by progressive augmentation. 2018.
33. Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
34. Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training. *Advances in Neural Information Processing Systems*, 33:7559–7570, 2020.