Why are Generative Adversarial Networks so Fascinating and Annoying?

Fabio Augusto faria
Institute of Science and Technology
Federal University of São Paulo - Brazil
Email: ffaria@unifesp.br

Gustavo Carneiro
Australian Institute for Machine Learning
The University of Adelaide - Australia
Email: gustavo.carneiro@adelaide.edu.au

***Abstract*—This paper focuses on one of the most fascinating and successful, but challenging generative models in the literature: the Generative Adversarial Networks (GAN). Recently, GAN has attracted much attention by the scientific community and the entertainment industry due to its effectiveness in generating complex and high-dimension data, which makes it a superior model for producing new samples, compared with other types of generative models. The traditional GAN (referred to as the Vanilla GAN) is composed of two neural networks, a generator and a discriminator, which are modeled using a minimax optimization. The generator creates samples to fool the discriminator that in turn tries to distinguish between the original and created samples. This optimization aims to train a model that can generate samples from the training set distribution. In addition to defining and explaining the Vanilla GAN and its main variations (e.g., DCGAN, WGAN, and SAGAN), this paper will present several applications that make GAN an extremely exciting method for the entertainment industry (e.g., style-transfer and image-to-image translation). Finally, the following measures to assess the quality of generated images are presented: Inception Search (IS), and Frechet Inception Distance (FID).**

I. Introduction

A generative model is represented by a probability distribution $p_{data}(x)$, where $x \in R^d$. Many generative models are learned by maximizing the likelihood of fitting the training data to a particular model parameterized by θ [1]. In particular, with independent and identically distributed (i.i.d.) training samples $\{x_i\}_{i=1}^n$, the likelihood is defined as the product of the probabilities that the model returns for each training data: $p_\theta(\{x_i\}_{i=1}^n) = \prod_{i=1}^n p_\theta(x_i)$, where $p_\theta(x)$ is the model probability attributed to x.

Figure 1 shows two categories of deep generative models that work by maximizing the likelihood: (a) explicit density models, which estimate $p_\theta(x)$ explicitly; and (b) implicit density models, which generate samples from $p_\theta(x)$, but cannot estimate this likelihood explicitly. GANs are classified as implicit density models and is the main topic of study of this paper.

Unlike traditional generative models, GANs are able to better represent complex data and learn high dimensional data distributions. Proposed by Goodfellow et al. [2], GANs are defined as a generative model optimised based on an adversarial training setup. In this approach, GANs are composed by two networks: a generator (G) and a discriminator (D). These networks are trained using a minimax optimization strategy,

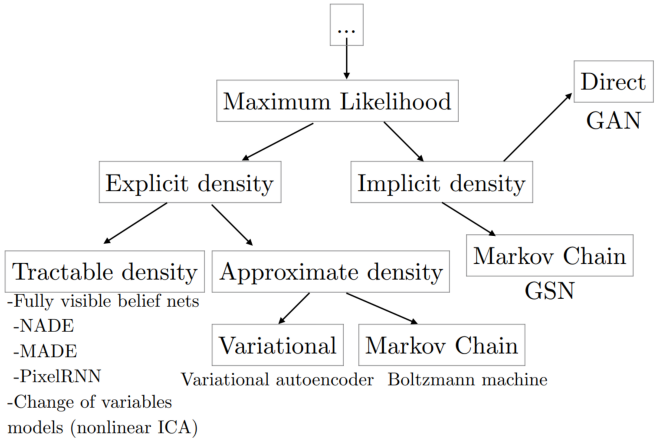

Fig. 1. Generative models models trained using maximum likelihood estimation. Image extracted from [1].

where an objective function is shared by both networks. On the one hand, the generator aims to fool the discriminator into believing that generated samples and real samples are from the same distribution. On the other hand, the discriminator aims to distinguish between generated (fake) and real samples [3].

Figure 2 shows a Vanilla GAN diagram, where generator network (G) has as input data a latent variable $z \in R^k$ sampled from a known distribution (Gaussian) and generates a fake data ($G(z)$). Discriminator network (D) aims to differentiate real and fake data.

GANs have become one of the most studied topics in machine learning and computer vision, a fact that can be evidenced by the increasing number of published papers about GANs and the fact that many researchers have adopted GAN-based approaches for several applications [5]. For instance, Figure 3 shows the cumulative number of published papers about GAN [1]. As it is possible to observe, in a four-year period the amount of papers increases from 2 in June/2014 to 502 papers in September/2018.

According to Goodfellow et al. [2], the Vanilla GAN approach had the following limitations which needed to be handled: (1) non-convergence; (2) mode collapse; (3) gradient uninformativeness; (4) overfitting; and (5) high sensitiveness.

[1] https://github.com/hindupuravinash/the-gan-zoo/blob/master/gans.tsv

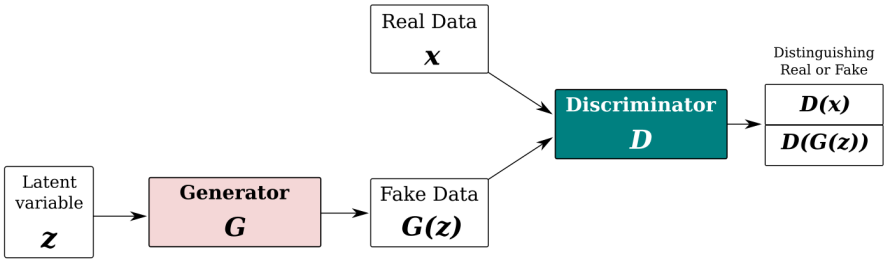

Fig. 2. The Vanilla GAN approach proposed in [4].

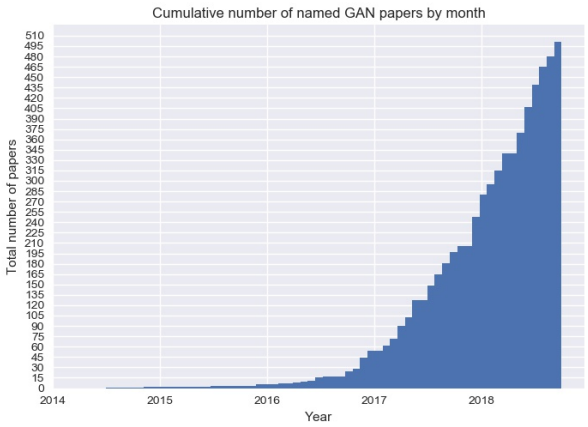

Fig. 3. Cumulative number of GAN papers.

Consequently, many papers in the literature have been proposed to deal with such limitations.

For instance, one way to stabilize training was to make changes to the objective function in order to avoid the vanishing gradient by clipping the gradient values and/or imposing penalties during training (e.g., Hinge [6], WGAN [7], and WGAN-GP [8]). Another way to stabilize training and postpone the discriminator's "victory" was proposed by Unrolled GAN [9]) by updating the generator parameters more often than updating the discriminator parameters.

In the line of architecture improvements, DCGAN [10] suggests the use of Convolutional Neural Networks (CNNs) with Batch Normalization, and ReLU activation. Progressive GAN [11] adopted a hierarchical structure of image generation. Other hybrid approaches have been proposed, combining autoencoder with GAN approaches (e.g., BEGAN [12] and MAGAN [13]). In addition, multiple generators might turn the minimax strategy a fairer game (MGAN [14], MAD-GAN [13], and MEGAN [15]). A set of papers engaged in modifying the layers of neural networks such as SN-GAN [16] and SAGAN [17], which proposed to use spectral normalization and attention concepts, respectively.

Another concern of current GAN research is the mode collapse problem, which is related to image diversity. Papers addressing this issue proposed keeping the distance in both latent and visual spaces, turning the approaches more "creative" (e.g., MSGAN [18] and DSGAN [19]). Regarding the quality of the images generated, some papers have sought to increase the spatial resolution of images up to 1024×1024 pixels (e.g., BigGAN [20], styleGAN [21] and MSG-GAN [22]).

With the emergence of conditional GANs (cGAN [23]), the most varied and creative applications have been proposed such as Image-to-Image (e.g., CycleGAN [24], pix2pix [25] , StarGAN [26], DIRT [27], starGANv2 [28], CartoonGAN [29]) and Text-to-Image (e.g., AttnGAN [30], StackGAN [31], ControlGAN [32], and StackGAN++ [33]).

Finally, several areas of knowledge have adopted GAN approaches to create training samples and improve their results in the target task (e.g., agriculture [34], biology [35], biometric [36], medicine [37], remote sensing [38], security [39]).

II. The Groundbreaking Approaches in Generative Adversarial Networks

In this section, we concentrate our explanations on the breakthroughs achieved by researchers in developing new GAN approaches.

A. *Wasserstein GAN (WGAN)*

Originally, GANs were designed to find a low value of a cost function using gradient descent techniques and they may fail to converge when used to seek for a Nash equilibrium. Hence, in [40], five techniques (feature matching, minibatch discrimination, historical averaging, one-sided label smoothing, and virtual batch normalization) have been proposed to improve the stability of training and the perceptual quality of generated samples.

In [7], Wasserstein GAN (WGAN) studied the behavior of the phenomenon called vanishing gradient in the generator when facing an optimal discriminator. Arjovsky et al. [7] showed that even when two distributions are located in lower dimensional manifolds without overlaps, the Wasserstein distance (Earth-Mover's distance) is a sensible cost function, unlike other measures, such as *Kullback-Leibler* divergence (KL), *Total Variation* distance (TV), and *Jensen-Shannon* divergence (JS). More specifically, the Wasserstein distance is a GAN loss function that provides a smooth measure, which is helpful for a stable learning process using gradient descent.

Furthermore, Arjovsky et al. [7] proposed a transformation of the formula based on the Kantorovich-Rubinstein duality, which involves a way to enforce a Lipschitz constraint through

weight clipping strategy. Thus, they measured the least upper bound (maximum value) of the Wasserstein distance and improved GAN performance.

B. Wasserstein GAN with Gradient Penalty (WGAN-GP)

The weight clipping strategy proposed for the WGAN approach works as a weight regularizer, limiting the ability of the generator to learn complex distribuitions. In this sense, WGAN-GP approach reduces the capacity of WGAN, making WGAN-GP training more stable. Table I shows the main objective functions proposed in the literature. It is important to note that WGAN and WGAN-GP distinguish each other by adding the gradient penalty at the end of the equation.

C. Spectral Normalization GAN (SN-GAN)

As mentioned before, Vanilla GAN is vulnerable to mode collapse and extremely unstable during training. SN-GAN tries to address these two problems simultaneously. A challenge that remains in GAN training is on how to control the learning of the discriminator. Therefore, in [16], a weight normalization technique called spectral normalization has been proposed to stabilize the training of the discriminator network. This normalization technique is concerned only with adjusting the Lipschitz constant, so there is no need to adjust the intensity of this hyper-parameter. More specifically, spectral normalization normalizes the weight for each layer with the spectral norm such that the Lipschitz constant for each layer as well as the whole network equals one.

D. Unrolled GAN

One of the major problems related to the mode collapse in Vanilla GANs is the quick learning of the discriminator in relation to the generator, resulting in a faster convergence for the discriminator in the minimax competition. The Unrolled GAN approach relies on a strategy based on a greater amount of updates of the weights/parameters for the generator network than for the discriminator. The motivation for this strategy is to let the generator predict the actions of the discriminator and postpone the mode collapse.

Figure 4 shows the behavior of Vanilla GAN compared to Unrolled GAN throughout many training epochs. As can be observed the Unrolled GAN (on top) can represent better the ground truth (target distribution) than the Vanilla GAN approach.

E. Self-Attention GAN (SAGAN)

As the first GAN approach to suggest the use of a convolutional neural network architecture, the DCGAN [10] adopted batch normalization and different types of activation functions to stabilize training process. Similarly, Self-Attention GAN (SAGAN) [17] proposed modifications to the Vanilla GAN architecture in order to improve the quality of images generated by the generator network. This new approach achieved surprising results in the literature with simple and efficient architecture called Transformer [41], which was an alternative to the complex deep and recurrent architectures (RNN and LSTM) through attention mechanism.

Figure 5 shows the proposed self-attention module for the SAGAN. For each architecture layer, the feature maps $x \in R^{C \times N}$ are transformed into three feature spaces $f(x) = W_f x$, $g(x) = W_g x$, and $h(x) = W_h x$. Next, $s_{ij} = f(x_i)^T g(x_j)$ and $\beta_{ji} = \frac{exp(s_{ij})}{\sum exp(s_{ij})}$ are computed, where β_{ji} indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region. Also, C is the number of channels and N is the number of feature locations of features from the previous hidden layer. Finally, the output of the attention layer $o_j = \beta_{ji} h(x_i)$ and final output $y_i = \gamma o_i + x_i$ are created, where γ is a learnable scalar and it is initialized as 0.

F. Conditional GANs (cGAN)

In [23], a conditional version of generative adversarial nets (cGANs) has been introduced through a simple way to feed the generator (G) and discriminator (D) networks with a conditional factor y. Figure 6 illustrates a general conditional GAN approach. This conditional factor y can represent different types of annotations (e.g., label, image, and text), and cGANs are applied in many different problems, such as:

- **Image synthesis by Label:** using coded label or one-hot vector as input data to create images [10], [18], [42], [43];
- **Image-to-image:** using paired images or unpaired images to create new images [24]–[28], [44];
- **Text-to-image:** using encoded text, embeddings or semantic attribute vector as input data to achieve new images [30]–[33].

Table I also shows the objective function of the cGAN.

G. Causality in GAN (CausalGAN)

CausalGAN [43] is an innovative approach that uses causality to control the generator through interventions and create images never seen in the set of real images. Figure 7 shows a causal graph, where each attribute/variable is represented by a node in this graph and the cause-effect relations between variables is represented by edges.

A causal controller module can use the causal graph to create new labels (L_g), which together with the noise vector z, represent the inputs of the generator network G and this network can generate an image that will be classified by the Discriminator D as real or fake. The role of the Labeler network, which has been trained with real images, is to estimate the labels of real images. The Anti-Labeler is trained with fake images by estimating their labels. In turn, the generator aims to minimize the Labeler loss (realistic images) and maximize the Anti-Labeler loss, stimulating the generator to create images that are different from those existing in the training dataset.

Figure 9 shows some examples of images generated through interventions and conditioning of distributions. Note that the intervention carried out on the $Male \rightarrow Bald$ relation resulted in the creation of images of bald women (on top), unlike the conditioned images, which show only bald men (on bottom).

TABLE I
EVOLUTION OF GAN OBJECTIVE FUNCTIONS.

GAN	Equation
Vanilla	$\min_G \max_D V(D,G) = E_{x\sim Pdata}[logD(x)] + E_{z\sim P(z)}[log(1-D(G(z)))]$
WGAN	$\min_G \max_D V(D,G) = E_{x\sim Pdata}[(x)] - E_{z\sim P(z)}[D(G(z))]$
Hinge	$\max_D V(D) = -E_{(x)\sim Pdata}[min(0,-1+D(x))] - E_{z\sim P(z)}[min(0,-1-D(G(z)))]$ $\min_G V(G) = -E_{z\sim P(z)}[D(G(z))]$
WGAN-GP	$\min_G \max_D V(D,G) = E_{x\sim Pdata}[(x)] - E_{z\sim P(z)}[D(G(z))] + \lambda\ E_{\hat{x}\sim P(\hat{x})}[(\parallel \nabla_{\hat{x}} D(\hat{x}) \parallel_2 -1)^2]$
cGAN	$\min_G \max_D V(D,G) = E_{x\sim Pdata(x)}[\ log\ D(x\|y)\ ] + E_{z\sim P(z)}[\ log(1-D(G(z\|y)))\ ]$

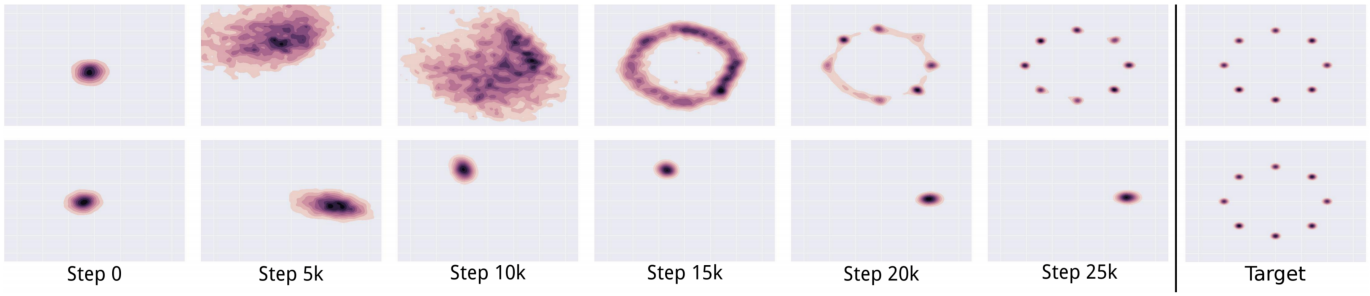

Fig. 4. Unrolled GAN (top) versus Vanilla GAN (bottom). Image extracted from [9].

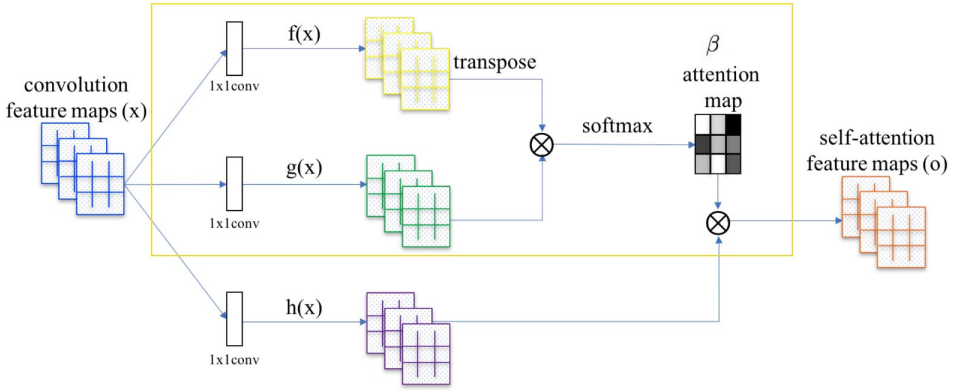

Fig. 5. The proposed self-attention module for the SAGAN. Image extracted from [41].

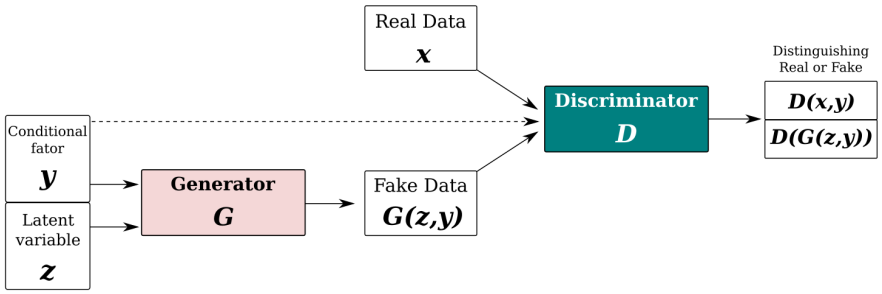

Fig. 6. A general cGAN approach.

H. Mode Seeking GAN (MSGAN)

The main idea of this conditional GAN approach is to address the mode collapse issue for cGANs through maximization of the ratio of the distance between generated images (visual space) and their corresponding latent codes (noise space), thus encouraging the generator to explore more minor modes during training [18]. The addition of a term in the generator loss function acts as a regularizer (mode seeking),

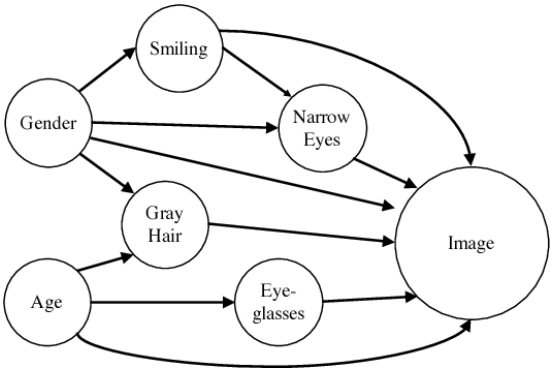

Fig. 7. Causal graph composed of six attributtes (Age, Eye-glasses, Gender, Gray Hair, Narrow Eyes, and Smiling) from CelebA dataset [45]. Image extracted from [43].

providing GAN with an extra dash of creativity/diversity in the image generation task.

Figure 10 shows examples of image generated by MSGAN and DIRT approaches. As can be observed, MSGAN obtains substantial diversity gains in comparison with DIRT.

III. Application Results

This section presents some well-known GAN-related applications proposed in the literature.

A. *Image-to-Image Translation*

Figure 11 shows some examples of unpaired image-to-image translation using CycleGAN [24]. This approach targets the translation between input and output images from two different domains (source $\rightarrow$ target). For instance, horse$\rightarrow$zebra and apple$\rightarrow$orange.

B. *Text-to-image Synthesis*

Figure 12 shows the evolution of the text-to-image synthesis in the literature. The rows show four different GAN approaches (GAN-INT-CLS [46], GAWWN [47], StackGAN-v1 [31], and StackGAN-v2 [33]) and in the columns are the corresponding images with the textual sentences. Notice that the StackGAN family approaches generate higher resolution images (256×256 pixels) with more photo-realistic details and diversity than the two other approaches in the literature.

C. *Evolution of Face Generation*

One of the most common image synthesis tasks is related to the creation of faces. This evolution of image creation starts with Vanilla GAN [4] using the TFD dataset [49] with 32 $times$32 pixels, passing through DCGAN [10] with images of 64×64 pixels and CoGAN [50] that generated images of 128×128 pixels. In 2018, ProGAN [11] was already able to generate images of 1024×1024 pixels through an architecture that incrementally increases the spatial resolution of the generated image. Next, NVIDIA proposed the styleGAN [21] and a new image dataset, called FFHQ. Currently, styleGAN, ProGAN with modified hyper-parameters and MSG:GAN [22] achieve state-of-the-art results in this problem.

Figure 13 shows nine different spatial resolutions of an image generated by multi-scale gradient approach, MSG:GAN. It is possible to observe the generative power of this approach, managing to generate images in dimensions of 1024×1024.

IV. Evaluation Measures

In literature, several measures have been proposed to assess the GAN performance. Two of the most employed are the Frechet Inception Distance (FID) and *Inception score* (IS) [51].

A. *Inception Score (IS)*

Inception Score (IS) measures the quality of an image generated by calculating the Kullback-Leibler divergence (D_{KL}) between the response (logit) produced by this image and the marginal distribution defined by $p(y) = \int_x p(y|x)\ p_g(x)$, i.e., the average response for all images generated using the model $p(y|x)$. In this case, this model is an Inception network pre-trained on the ImageNet dataset. Then, IS is computed with:

$$IS(p_g) = exp(\ E_{x \sim p_g}[\ D_{KL}\ (\ p(y|x)\ \|\ p(y)\ )\ ]\ ) \quad (1)$$

B. ***Fréchet Inception Distance (FID)***

FID compares the activations of the pre-trained Inception network (penultimate layer) among sets of real (x) and generated (g) images. This comparison approximates those sets with Gaussian distributions, calculating their means (μ) and covariances (Σ) [51].

Given the features from the x and g images, in this case a 2048-dimensional activation of the third layer pooling of the Inception-v3 network, FID compares the means μ_x and μ_g as well as the covariances Σ_x and Σ_g of the features:

$$FID(x, g) = \|\mu_x - \mu_g\|_2^2 + Tr(\Sigma_x + \Sigma_g\ -\ 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}). \quad (2)$$

FID metric is sensitive to mode collapse and is more robust to noise than IS.

Although IS and FID metrics are the most used GAN assessment measures in literature, many other works have proposed new metrics that can measure the quality of GAN representations. A critical point not captured by these two measures [52] is their inability to assess the relevance of generated images to the learning process of CNNs. Therefore, GAN-Train and GAN-Test metrics have been proposed to overcome that problems.

V. By the way, why are they "annoying"?

As this paper was written during a one-year sabbatical leave, it describes the author's experience about target subject. Of course, the term "annoying" is a bit strong and supposed to be a joke about GAN approaches. Although GAN approaches provide enormous power to represent complex and high dimensional distributions, working with them is not a trivial task for beginners. This section aims to show some important points that should be looked more carefully as well as some tips.

1) Many published works can be a negative factor when someone wants to learn about GANs. Tutorials like this

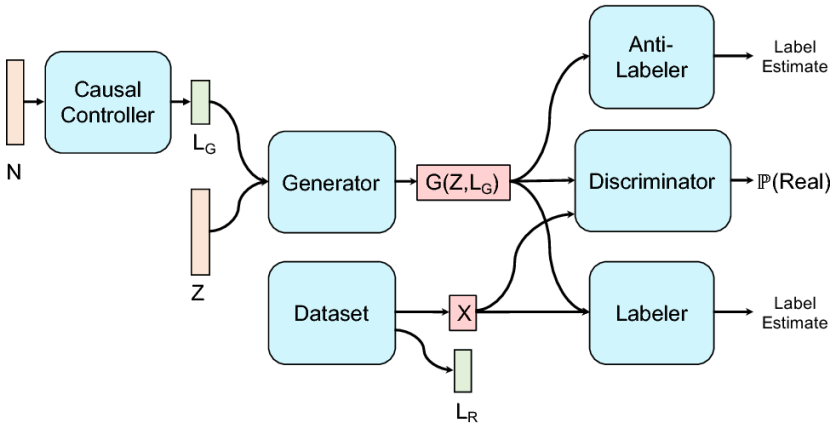

Fig. 8. CausalGAN diagram. Image extracted from [43].

Fig. 9. Examples of generated images through interventions. Top: Intervened on $Bald = 1$. Bottom: Conditioned on $Bald = 1$. Image extracted from [43].

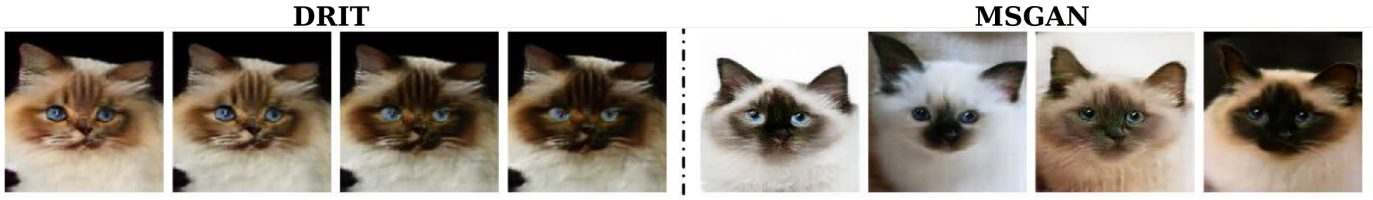

Fig. 10. Examples of images generated by DIRT and MSGAN approaches. Image adapted from [18].

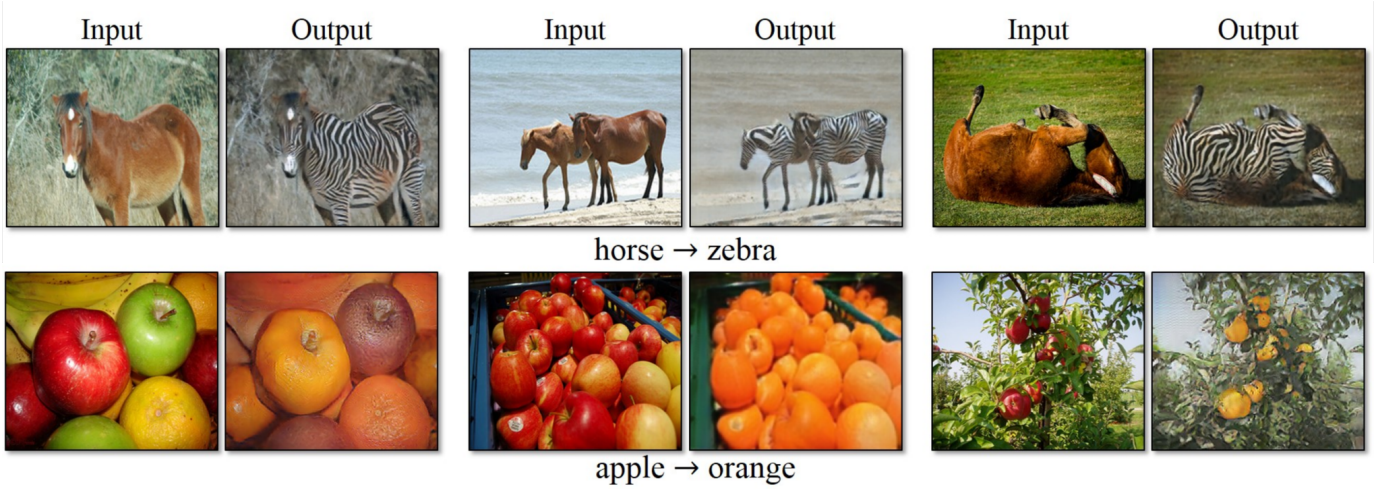

Fig. 11. Examples of mapping $source \rightarrow target$ using CycleGAN. Image adapted from [24].

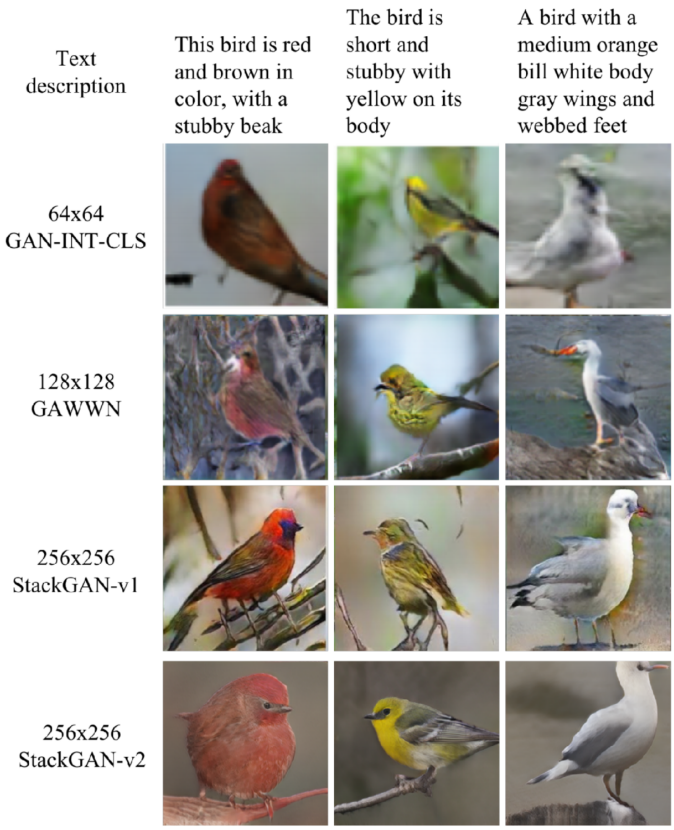

Fig. 12. Example results by StackGAN family approaches, GAWWN, and GAN-INT-CLS conditioned on text descriptions from CUB testset [48]. Image extracted from [33].

help and they are good alternatives for improving the understanding of GANs;

2) Several new and/or complex mathematical concepts are commonly used in this area and they need to be learned by reader;
3) In typical GAN training, the discriminator training converges faster than the generator training, so the discriminator can classify if the image is real or fake much faster than the generator can produce deceivable images;
4) Difficulty in stabilizing the learning. A recommendation is to make small and gradual changes along the GAN pipeline (e.g., loss function, architecture, layers, image resolution and application);
5) "No free lunch theorem" is very present in this area. There is not a generic GAN to perform any kind of application or even for the same application with different datasets. Therefore, for any simple change, probably some adjustment might be needed for the hyper-parameters and architecture;
6) In terms of data augmentation (DA), although DA techniques based on geometric transformation (GT) are visually less compelling than GAN-based samples, for learning a classifier, the GT-based samples might be more effective.
7) Training time does not mean better images. It is recommended to look at the loss function curves (generator and discriminator) as well as the quality of the generated images at each training epoch;
8) GANs can create visually pleasing images, but there is no guarantee that these images can contribute to training CNNs. The closer to the real distribution (low FID), the less relevant the generated images can be as training samples;

VI. Acknowledgments

The authors thank the support of the São Paulo Research Foundation (FAPESP) through grants #2017/25908-6 and #2018/23908-1. Also the Brazilian scientific funding agency CNPq through the Universal Project (grant #408919/2016-7).

References

[1] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," 12 2016.
[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds., 2014, pp. 2672–2680.
[3] H. Zhang, Z. Zhang, A. Odena, and H. Lee, "Consistency regularization for generative adversarial networks," in *International Conference on Learning Representations – to appear*, 2020.
[4] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *CoRR*, vol. abs/1412.6572, 2014.
[5] Y. Hong, U. Hwang, J. Yoo, and S. Yoon, "How generative adversarial networks and their variants work," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–43, 2019.
[6] J. H. Lim and J. C. Ye, "Geometric gan," 2017.
[7] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International Conference on Machine Learning*, vol. 70, 2017, pp. 214–223.
[8] L. Castrejón, N. Ballas, and A. C. Courville, "Improved conditional vrnns for video prediction," in *2019 IEEE/CVF International Conference on Computer Vision*. IEEE, 2019, pp. 7607–7616.
[9] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks," in *International Conference on Learning Representations*, 2017.
[10] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *International Conference on Learning Representations*, 2016.
[11] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *International Conference on Learning Representations*, 2018.
[12] Y. Li, N. Xiao, and W. Ouyang, "Improved boundary equilibrium generative adversarial networks," *IEEE Access*, vol. 6, pp. 11 342–11 348, 2018.
[13] A. Ghosh, V. Kulharia, V. P. Namboodiri, P. H. S. Torr, and P. K. Dokania, "Multi-agent diverse generative adversarial networks," in *Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2018, pp. 8513–8521.
[14] Q. Hoang, T. D. Nguyen, T. Le, and D. Q. Phung, "MGAN: training generative adversarial nets with multiple generators," in *International Conference on Learning Representations*, 2018.
[15] Y. Sun, S. Wang, T.-Y. Hsieh, X. Tang, and V. Honavar, "Megan: A generative adversarial network for multi-view network embedding," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 7 2019, pp. 3527–3533. [Online]. Available: https://doi.org/10.24963/ijcai.2019/489
[16] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *International Conference on Learning Representations*, 2018.

Fig. 13. Nine different spatial resolutions of face images generated by MSG:GAN at epoch 230. Image adapted from [22].

[17] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97, 2019, pp. 7354–7363.

[18] Q. Mao, H. Lee, H. Tseng, S. Ma, and M. Yang, "Mode seeking generative adversarial networks for diverse image synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1429–1437.

[19] D. Yang, S. Hong, Y. Jang, T. Zhao, and H. Lee, "Diversity-sensitive conditional generative adversarial networks," in *International Conference on Learning Representations*, 2019.

[20] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *International Conference on Learning Representations*, 2019.

[21] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition, pages = 4401–4410, year = 2019,*.

[22] A. Karnewar, O. Wang, and R. S. Iyengar, "MSG-GAN: multi-scale gradient GAN for stable image synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition, year = 2020,*.

[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *International Conference on Computer Vision*, 2017.

[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[26] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

[27] H. Lee, H. Tseng, J. Huang, M. Singh, and M. Yang, "Diverse image-to-image translation via disentangled representations," in *European Conference Computer Vision*, vol. 11205, 2018, pp. 36–52.

[28] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[29] Y. Chen, Y.-K. Lai, and Y.-J. Liu, "Cartoongan: Generative adversarial networks for photo cartoonization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[30] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," 2018.

[31] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks."

[32] B. Li, X. Qi, T. Lukasiewicz, and P. H. S. Torr, "Controllable text-to-image generation," 2019.

[33] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan++: Realistic image synthesis with stacked generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1947–1962, 2019.

[34] R. Barth, J. Hemming, and E. Van Henten, "Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation," *Computers and Electronics in Agriculture*, vol. 173, p. 105378, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169919320794

[35] A. Osokin, A. Chessel, R. E. C. Salas, and F. Vaggi, "GANs for Biological Image Synthesis," in *ICCV 2017 - IEEE International Conference on Computer Vision*, Venice, Italy, Oct. 2017. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01611692

[36] J. Zhang, Z. Lu, M. Li, and H. Wu, "Gan-based image augmentation for finger-vein biometric recognition," *IEEE Access*, vol. 7, pp. 183 118–183 132, 2019.

[37] A. Bissoto, F. Perez, E. Valle, and S. Avila, "Skin lesion synthesis with generative adversarial networks," in *Workshop MICCAI*, vol. 11041, 2018, pp. 294–302.

[38] Y. Yu, X. Li, and F. Liu, "Attention gans: Unsupervised deep feature learning for aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 519–531, 2020.

[39] J. Jenkins, K. Roy, and J. Shelton.

[40] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems, pages = 5998–6008, year = 2017,*.

[42] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, p. 2180–2188.

[43] M. Kocaoglu, C. Snyder, A. G. Dimakis, and S. Vishwanath, "CausalGAN: Learning causal implicit generative models with adversarial training," in *International Conference on Learning Representations*, 2018.

[44] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 465–476.

[45] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[46] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative Adversarial Text to Image Synthesis," ser. Proceedings of Machine Learning Research, vol. 48, 20–22 Jun 2016, pp. 1060–1069.

[47] S. E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, "Learning what and where to draw," 2016.

[48] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, "Caltech-UCSD Birds 200," California Institute of Technology, Tech. Rep. CNS-TR-2010-001, 2010.

[49] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling factors of variation for facial expression recognition," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI*. Berlin, Heidelberg: Springer-Verlag, 2012, p. 808–822.

[50] M.-Y. Liu and O. Tuzel, "Coupled Generative Adversarial Networks," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 469–477. [Online]. Available: http://papers.nips.cc/paper/6544-coupled-generative-adversarial-networks.pdf

[51] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, 2017, pp. 6626–6637.

[52] K. Shmelkov, C. Schmid, and K. Alahari, "How Good Is My GAN?" in *European Conference on Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., 2018, p. 218–234.