Synthetic Images: Generative Adversarial Networks and Diffusion

CS4991 Capstone Report, 2023

Luke Benham
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
lnb6grp@virginia.edu

ABSTRACT
The trend in the development of Artificial Intelligence is towards machine learning involving deep neural networks that are increasingly opaque to human understanding. Generative Adversarial Networks exemplify this issue as they use unsupervised training and the discriminator model iteratively creates the utility function for the generation model. Diffusion extends this issue as the training involves reversing random Gaussian noise which leads to a level of randomness that is near impossible to decode once the model is trained. The lack of transparency leads to issues detecting bias, bugs, or unintentional outputs from the models. Despite these potential downsides, image generators have demonstrated incredible abilities that will transform the online landscape.

1. INTRODUCTION

While the intent of an artist may influence the understanding of a particular work and be debated among art critics, the intent of Artificial Intelligence systems will be of societal importance. The legibility and alignment of AI systems has become more difficult to discern as more complex systems are developed, with image generators being some of the most inscrutable. If we cannot understand the ways AI can create art, then how can we ensure these systems can be trusted with more powers over society?

The use of computers to create art goes back over 50 years, with the first significant example being Cohen's AARON system which was designed to mimic the act of drawing. Art creating programs remained a primarily rule-based and procedural until the development of deep neural networks were able to compute and understand images at the pixel level. This began with the development of LeNet in 1998 which used a convolutional neural network to decrease the dimensions of an image at each step to a classification of a handwritten character (Lecun et al., 1998). Image classification AIs continued to improve upon the use of convolutional networks and gradient backpropagation to increase the complexity and accuracy of these systems. These advancements led to the creation of the Generative Adversarial Network (Goodfellow et al., 2014).

2. RELATED WORKS

Hong, et al., 2020 provides a survey of techniques and details used in the creation of GANs. This work both describes the basics for GAN development but also the potential uses in fields such as "image synthesis, image attribute editing, image translation, domain adaptation, and other academic fields."

Luo, 2021 also surveys the field of image synthesis with a focus on "multimodal deep learning image generation." This field tackles the challenge of teaching models to understand the connections between various

forms of data, including text, images, video and more. This is a more difficult problem than traditional GANs and shows the potential for development in new and challenging fields.

3. META-STUDY

This meta-study will examine the two most significant techniques in artificial image generation. While GANs were followed by the development of diffusion-based models, both are often used in conjunction to create cutting-edge models.

3.1. Generative Adversarial Networks

Adversarial networks are based on the creation of two competing systems that are used to train and improve based on the output of the opposing system. Generally, only one of the systems will be the desired model and the other is used solely in training. In the case of GANs these are the generator and discriminator models with the generator being the desired model (Figure 1 below).

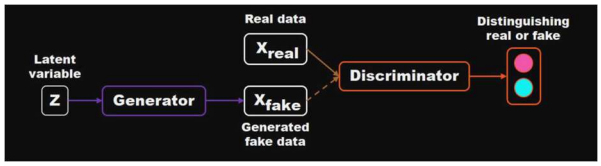

Figure 1: Diagram of the structure of a Generative Adversarial Network model. (Hong, et al., 2020, p. 4)

The generator model takes a batch of latent seed variables to avoid a deterministic output and creates a series of synthetic images. These are fed to the discriminator along with real images which output a scalar value describing the probability of a generated image. The discriminator then updates based on the equation in Figure 2.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$$

Figure 2: Equation for stochastic gradient ascent of the discriminator model. (Goodfellow et al., 2014, p.4)

The discriminator is trained for k steps before the samples are used to update the generator model according to the equation in Figure 3.

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

Figure 3: Equation for stochastic gradient descent of the generator model. (Goodfellow, et al., 2014, p.4)

This design as described by Goodfellow, et al. in 2014 forms the basis for GANs as they have been created up until the current day. Methods have been refined and adapted as GANs are used in other fields such as video GANs and multimodal models.

3.2 Diffusion

The first diffusion models in machine learning were created as an analogy to non-equilibrium thermodynamics (Sohl-Dickstein, et al., 2015). The data would be degraded through increasing entropy or noise like the diffusion of a gas, but unlike the physical version the neural network would be trained to reverse this process. These early models were trained by adding noise to an image and having the model reverse the process to the original image. Then, when prompted with a random field of noise the model could be prompted to generate a new image.

The leap forward in the abilities of diffusion models came from the transition to training over the noised pixels to training over an abstracted latent space (Rombach, et al., 2021). The latent space is a compressed high dimensional version of an image that retains the semantic meaning of the image at a smaller scale. By using this latent space, the model is able to perform the diffusion process on a more abstracted form of the image as seen in Figure 4.

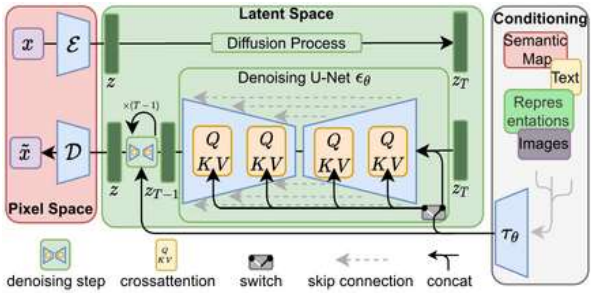

Figure 4: Process for training a diffusion model using latent space. (Rombach, et al., 2021)

After these advances diffusion models began outperforming GANs in the quality of images produced, as well as having advantages with other techniques. While GANs are optimized for creating images that are indistinguishable from the real images in the training set, the inherent randomness of the diffusion training leads to increased flexibility in model usage. Diffusion-based models show greater capabilities in image-to-image translation by noising a seed image and prompting to transition to a partially generated image (Saharia et al., 2022). A more basic and direct avenue for their advantages is in effectively de-noising images through the diffusion process, though this may need to be augmented to avoid artifacts created by the model (Qiao, et al., 2017).

4. OUTCOMES

The development of generative adversarial networks and diffusion models is a demonstration of how advancements in artificial intelligence often occur. The field shows certain trends in capabilities, often correlated with processing power, but certain seminal papers and paradigm shifts create leaps forward in the power of these models.

5. CONCLUSION

The continued development of GANs and diffusion models will expand the visual creativity available to humanity. The barrier to entry for artistic creation will continue to fall as areas previously reserved for human artists are reached by AIs. Adversarial networks will see continued usage, and the discriminator model may become a goal instead of purely a training step to be used to determine true from synthetic images. Diffusion models will improve both in realism and creativity. Multimodal models will likely be an expanding field, with diffusion models using diverse types of prompts to better match the vision of the users. Whether AI-generated images will be considered art is unclear, but the rights attributed to AI systems will become a notable issue.

6. FUTURE WORK

As the field of image generation is continuously advancing the space for new meta-studies will expand along with frontier capabilities. Image models are increasingly being used in multimodal models which offers avenues of investigation into a developing and complex field. Also image synthesis may move from more static models to AI agents with greater abilities due to self-prompting and work outside of human input.

REFERENCES

Aldausari, N., Sowmya, A., Marcus, N., & Mohammadi, G. (2022). Video generative adversarial networks: a review. *ACM Computing Surveys, 55*(2), 1–25. https://doi.org/10.1145/3487891

Barnett, M. (2020, August 23). *When will the first general AI system be devised, tested, and publicly announced?* Date of Artificial General Intelligence. Retrieved November 1, 2022, from https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/

Canas, K., Ubiera, B., Liu, X., & Liu, Y. (2018). Scalable biomedical image

synthesis with gan. *Proceedings of the Practice and Experience on Advanced Research Computing,* 1–3. https://doi.org/10.1145/3219104.3229261

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., &; Bengio, Y. (2014). Generative Adversarial Networks. Communications of the ACM, 63(11), 139–144. https://doi.org/10.1145/3422622

Hong, Y., Hwang, U., Yoo, J., & Yoon, S. (2020). How generative adversarial networks and their variants work. *ACM Computing Surveys, 52(1),* 1–43. https://doi.org/10.1145/3301282

Lecun, Y., Bottou, L., Bengio, Y., &; Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Luo, S. (2021). A survey on multimodal deep learning for image synthesis. *The 5th International Conference on Innovation in Artificial Intelligence,* 108–120. https://doi.org/10.1145/3461353.3461388

Qiao, P., Dou, Y., Feng, W., Li, R., & Chen, Y. (2017). Learning non-local image diffusion for image denoising. *Proceedings of the 25th ACM International Conference on Multimedia,* 1847–1855. https://doi.org/10.1145/3123266.3123370

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., & Norouzi, M. (2022). Palette: Image-to-image diffusion models. *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings,* 1–10. https://doi.org/10.1145/3528233.3530757

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. &; Ganguli, S.. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. Proceedings of the 32nd International Conference on Machine Learning, in Proceedings of Machine Learning Research 37:2256-2265. https://proceedings.mlr.press/v37/sohl-dickstein15.html.