# Why Diffusion Models

**Shengding Ni** [1]

## Abstract

Diffusion models have emerged as a new family of deep generative models with unprecedented performances. By comparing diffusion models with other mainstream generative models in terms of stability, sampling quality and computational resources, we demonstrate the edge of diffusion models. This survey aims to provide an in-depth look at the current state of diffusion models, indicating why, where and how to use diffusion models.

*Figure 1.* Ink diffuses in water like data diffuse from a low-dimensional manifold to a high-dimensional normal distribution.

## 1. Introduction

Generative Models aim to approximate the probability distribution of the data and to generate similar ones, through which way they promise to endow machines with an understanding of the real world.

In the past decade, mainstream generative models including GAN (Goodfellow et al., 2014), VAE (Kingma & Welling, 2013), Flow (Dinh et al., 2014) have achieved great success in the field of computer vision (Brock et al., 2018; Razavi et al., 2019; Ho et al., 2019). Now, diffusion models (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020) have emerged as the new state-of-the-art family of deep generative models, rivaling GAN in the challenging task of image generation (Dhariwal & Nichol, 2021), natural language processing (Austin et al., 2021) and multi-modal modeling (Saharia et al., 2022; Avrahami et al., 2021; Rombach et al., 2021). Despite the achievements, research on diffusion models is in its early stages (Yang et al., 2022), with much potential for improvement in both theoretical and empirical aspects.

In this study, we first briefly review the development of diffusion models and its mechanism (Sec.2). Following that, we provide an overview of diffsion models applicatons and current major improvements.

[1]Department of Physics, Fudan University, Shanghai, China. Correspondence to: Shengding Ni <21307110025@m.fudan.edu.cn>.

*Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. 

Next, we compare diffusion models with other mainstream generative models in Sec.3. Without employing complicated mathematics, we interpret different generative models in a certain level of abstraction, and based on this, we analyze the pros and cons of each model in specific scenarios.

Finally, in Sec.4, we fine tune a diffusion model to synthesis pictures of specific style. Experimentally, we bump up against problems facing many diffusion model learner, and we conquer them utilizing existing techniques (Nichol & Dhariwal, 2021; Song et al., 2020; Dhariwal & Nichol, 2021; Ho et al., 2021; Ho & Salimans, 2022).

Key contributions of the paper include:

1. There are currently few reviews on diffusion models, yet the body of research on diffusion models has grown significantly in the past few years. Our research can make up for the lack in this area.

2. We conduct an in-depth study on relationships as well as comparisons between various generative models, which previous research rarely paid attention to.

3. We provide valuable suggestions for beginners of diffusion models, demonstrating why they should employ diffusion models, what improvements can be made, and in what scenarios they should apply diffusion models.

## 2. What are Diffusion Models

What does 'diffusion' refers to in the context of diffusion model? Imagine that ink diffuses in water because of thermal motion (as shown in fig.1). This is diffusion, the process in which data points diffuse from a low-dimensional manifold to a high-dimensional normal distribution. Only in the space of probability can we 'see' this wonderful process. Here, we unfold diffusion models in terms of development, mechanisms, applications, and improvements.

### 2.1. Development

Diffusion probabilistic models was first proposed in 2015 (Sohl-Dickstein et al., 2015), but it has not been widely used until the noise-conditioned score network (Song & Ermon, 2019) and denoising diffusion probabilistic models (Ho et al., 2020) came out. Their work improved the architecture of the model, making it easier to train, with better sampling quality.

After that, huge amounts of methods have been developed to improve diffusion mdoels with respect to sampling methods (Nichol & Dhariwal, 2021; Song et al., 2020), log-likelihood (Nichol & Dhariwal, 2021), classifier guidance (Dhariwal & Nichol, 2021; Ho & Salimans, 2022; Ho et al., 2021) and etc..

However, research on diffusion models is in its early stages (Yang et al., 2022), with much potential for improvements in both theoretical and empirical aspects.

### 2.2. Mechanism

We briefly review the mechanism of DDPMs (Ho et al., 2020).

**Diffusion process and Reverse process:** Given a data distribution $x_0 \sim q\left(x_0\right)$, we define a schotastic process $q$ that gradually add noises to the data as follows:

$$q\left(x_{1:T}|x_0\right) := \prod_{t=1}^{T} q\left(x_t|x_{t-1}\right) \tag{1}$$

$$q\left(x_t|x_{t-1}\right) := \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \mathbb{I}\right) \tag{2}$$

Given sufficiently large $T$ and a well behaved schedule of $\beta_t$, $x_T$ is nearly a Gaussian distribution. Therefore, we can sample $x_T \sim \mathcal{N}\left(0, \mathbb{I}\right)$ and gradually restore the data by reversing the diffusion process:

$$q\left(x_{t-1}|x_t, x_0\right) = \mathcal{N}\left(x_{t-1}; \tilde{\mu}_t\left(x_t, x_0\right), \tilde{\beta}_t \mathbb{I}\right) \tag{3}$$

$$\tilde{\mu}_t\left(x_t, x_0\right) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}x_t \tag{4}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{5}$$

However, the process rely on $x_0$. So we approximate $q\left(x_{t-1}|x_t, x_0\right)$ using a neural network:

$$p\left(x_T\right) = \mathcal{N}\left(x_T; 0, \mathbb{I}\right) \tag{6}$$

$$p_\theta\left(x_{t-1}|x_t\right) := \mathcal{N}\left(x_{t-1}; \mu_\theta\left(x_t, t\right), \Sigma_\theta\left(x_t, t\right)\right) \tag{7}$$

With a combination of diffusion and reverse process, we can calculate the variational lower bound of $-\log p_\theta\left(x_0\right)$ as follow:

$$L := L_0 + L_1 + \cdots + L_T \leq -\log p_\theta\left(x_0\right) \tag{8}$$

$$L_0 := -\log p_\theta\left(x_0|x_1\right) \tag{9}$$

$$L_{t-1} := D_{KL}\left(q\left(x_{t-1}|x_t, x_0\right)||p_\theta\left(x_{t-1}|x_t\right)\right) \tag{10}$$

$$L_T := D_{KL}\left(q\left(x_T|x_0\right)||p_\theta\left(x_T\right)\right) \tag{11}$$

**Simplification:** According to Eq.1 and Eq.2, and through reparameterization, we can write:

$$q\left(x_t|x_0\right) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, \left(1-\bar{\alpha}_t\right)\mathbb{I}\right) \tag{12}$$

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{13}$$

where $\alpha_t := 1-\beta_t$, $\bar{\alpha}_t := \prod_{s=1}^{t}\alpha_s$ and $\epsilon \sim \mathcal{N}\left(0, \mathbb{I}\right)$.

According to Eq.3 and Eq.13, we can randomly sample $t$ and use the expectation $E_{t,x_0,\epsilon}\left[L_{t-1}\right]$ to estimate $L$ in Eq.8.

Additionally, rewrite $\mu_\theta\left(x_t, t\right)$ and simplify $L$ as follows:

$$\mu_\theta\left(x_t, t\right) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta\left(x_t, t\right)\right) \tag{14}$$

$$L_{simple}\left(\theta\right) := E_{t,x_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right] \tag{15}$$

It proves that optimizing this reweighted objective resulted in much better sample quality than optimizing $L$ directly (Ho et al., 2020; Song & Ermon, 2019).

**Training:** In practice, the schedule of $\beta_t$ can be linear (Ho et al., 2020) or cosine (Nichol & Dhariwal, 2021). The parameters of variance $\Sigma_\theta\left(x_t, t\right)$ are fixed as $\sigma_t^2\mathbb{I}$, with $\sigma_t^2 = \beta_t$ or $\sigma_t^2 = \tilde{\beta}_t$.

Then, the simplified objective $L_{simple}$ in Eq.15 is optimized through back propagations.

### 2.3. Applications

Diffusion models have recently been employed to address a variety of challenging real-world tasks due to their flexibility and strength.

**Computer Vision:** Diffusion models have been used to tackle image restoration tasks including super-resolution and inpainting. For example, super resolution via Repeated Refinement (Saharia et al., 2023) uses DDPM to enable conditional image generation, and RePaint (Lugmayr et al.,

2022) use denoising strategy based on diffusion models to inpaint shaded or broken image.

**Natural Language Generation:** Besides generating images of high quality, diffusion models can also be used for text generation because of its flexibility. For example, Discrete Denoising Diffusion Probabilistic Models (D3PM) (Austin et al., 2021) introduces modified diffusion models for text generation.

**Multi-Modal Generation:** One of the most impressive applications of diffusion models is Text-to-Image generation (Saharia et al., 2022), where images are generated according to a descriptive text. For example, Blended diffusion (Avrahami et al., 2021) utilizes both pre-trained DDPM and CLIP models.

### 2.4. Improvments

A variety of methods have been developed, either to improve performances of diffusion models or to mitigate the cost of training and sampling processes.

**Efficient Sampling:** Generating samples from diffusion models typically rely on a long Markov chain that requires much time. Recent improvments, such as strided sampling schedule (Nichol & Dhariwal, 2021) and DDIM (Song et al., 2020), speed up the sampling process while also improving quality of the resulting samples.

**Improved Likelihood:** The training objective for diffusion models is the variational lower bound of the log-likelihood, which is not necessarily tight in many cases. By optimizing the noise schedule $\beta_t$ (from linear to cosine), one can approach and even saturate the bound (Nichol & Dhariwal, 2021).

**Classifier Guidance:** With conditioning information such as class labels of images, ADM and ADM-G (Dhariwal & Nichol, 2021) train a classifier on noisy images and use guidance from a classifier to run conditioned generation, which are able to achieve better results than SOTA generative models. Furthermore, CDM (Ho et al., 2021) employs noise conditioning augmentation to scale up generation resolution and quality.

## 3. Connections with other Generative Models

Generative models vary in their loss functions. However, none of them is perfect metrics of probability distribution. Therefore, maybe in the future a better model will emerge that beats all existing models. But now, we have to balance the pros and cons of each model and use them in suitable scenarios. Here, we briefly review mainstream generative models as shown in fig.2. By comparing with other models, we can clearly see the edge of diffusion models.

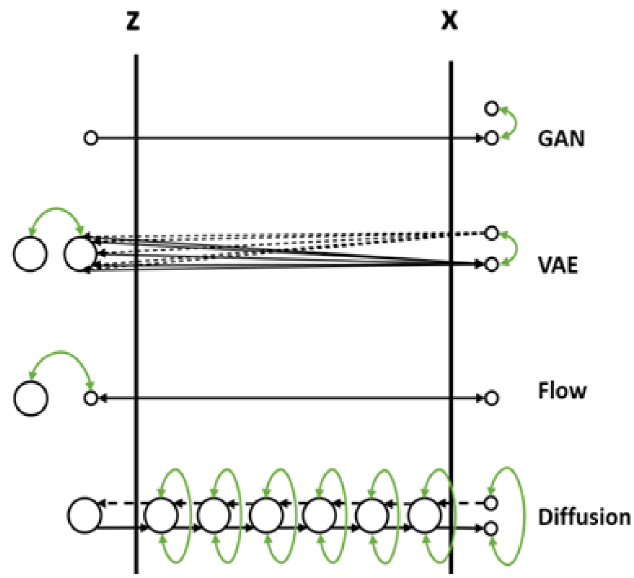


*Figure 2.* Overview of different types of generative models. The small circles are data points, and the big circles are data distributions. Black arrows represent neural networks (the solid line contains the parameters to be learned, while the dotted line is usually fixed), and green lines represent certain loss functions

### 3.1. Generative Adversarial Nets

GAN (Goodfellow et al., 2014) consists of a generator $G(\cdot)$ and a discriminator $D(\cdot)$, with the generator attempting to produce data $x$ that can fool the discriminator, and the discriminator learning to correctly identify synthetic data. Two networks are trained simultaneously. (SOTA: BigGAN-deep (Brock et al., 2018))

**Applications:** Image Generation (Brock et al., 2018; Karras et al., 2019), Anomaly detection (Schlegl et al., 2017).

**Pros:** GAN does not rely on Markov chains and variational inference, on which VAE and DDPMs are based; GAN can generate clear picture with high fidelity.

**Cons:** GAN suffers from unstable training since Nash equilibrium is not necessarily achieved (Salimans et al., 2016); The gradients of the loss function tend to vanish as discriminator performs better (Arjovsky & Bottou, 2017); The generator may collapse and always produces same outputs, referred to as Mode Collapse (Arjovsky et al., 2017).

**Summary:** GAN can quickly generate clear image. But its training is unstable and suffer from mode collapse and lack of diversity. You can use GAN if you have demands on speed and clearness of image generation.

### 3.2. Variational Autoencoder

VAE (Kingma & Welling, 2013) follows the idea of autoencoder, but the input $x$ is no longer mapped to a fixed latent variable $z$, but a probability distribution. The lower bound of Log-likelihood is aquired through variational inference, referred as ELBO. (SOTA: VQ-VAE-2 (Razavi et al., 2019))

**Applications:** Image Generation (Razavi et al., 2019), Image Processing (Rombach et al., 2021).

**Pros:** The architecture of VAE is clear and straighforward, making it easier and more stable to train.

**Cons:** Images generated by VAE are vague compared to that of GAN, mostly because ELBO is optimized rather than Log-likelihood itself.

**Summary:** VAE, boasting its simple and universal structure, has broader applications. However, sampling quality is a fatal problem. You can use VAE with requirement on stability. Also, VAE is useful when it comes to Image-to-Image translation.

### 3.3. Flow-based Deep Generative Models

A flow-based generative model (Dinh et al., 2014) consists of a sequence of invertible transformations. Unlike GAN and VAE, the model explicitly learns the data distribution and therefore the loss function is simply the negative Log-likelihood. (SOTA: Flow++ (Ho et al., 2019))

**Applications:** Image Generation (Ho et al., 2019), Anomaly Detection (Rudolph et al., 2020).

**Pros:** There is nothing more exciting than explicitly learning the data distribution. The flow model, based on Normalizing flows, is completely tractable in learning, sampling and estimating; Normalizing flows are relatively easy to implement, compared to GAN and VAE.

**Cons:** Flow models can be computationally expensive to train, especially when dealing with high-dimensional data; The training can be unstable when learning complex distribution compared to VAE; Poor sampling quality.

**Summary:** Flow models achieve tractability at cost of flexibility and computational resources. You can use flow model if you need explicit expression of probability distribution.

### 3.4. Denoising Diffusion Probabilistic Models

The DDPM can be conceptualized as a hierarchical Markovian VAE. In other words, each step of the DDPM can be regarded as a tiny VAE.

**Applications:** Image Generation (Avrahami et al., 2021; Saharia et al., 2022; Rombach et al., 2021), Text Generation (Austin et al., 2021).

*Table 1.* Camparisons between various generative models in general cases. The more star, the better performance.

| MODELS | QUALITY | RESOURCES | STABILTIY |
|---|---|---|---|
| GAN | ★★ | ★★★ | ★ |
| VAE | ★ | ★★★ | ★★★ |
| FLOW | ★ | ★ | ★★ |
| DDPMS | ★★★ | ★ | ★★★ |

**Pros:** DDPMs simultaneously achieve both flexibility and tractability(Sohl-Dickstein et al., 2015); DDPMs are readily applicable to a wide range of senarios with excellent sampling quality.

**Cons:** Diffusion models rely on a long Markov chain of diffusion steps to generate samples, so it can be quite expensive in terms of time and compute. New methods have been proposed to make the process much faster, but the sampling is still slower than GAN (Weng, 2021).

**Summary:** DDPMs exchange time and space for tractability and flexibility, which is worthwhile. In most cases, people care most about the quality of outcomes, which explains why there are so many research and applications of diffusion models.

Here we roughly evaluating performances of various generative models. The result is presented in Table 1.

## 4. Experiment

Experimentally, we reproduce functions of DDPMs (Ho et al., 2020) that run with one click on jupyter notebook and verify it on the CIFAR10 dataset (Krizhevsky et al., 2009). Furthermore, we train DDPMs on oil painting dataset and anime face dataset to generate images of different styles.

### 4.1. Details

There exists a few points you should pay attention to when training a diffusion model.

**Timesteps and noise schedule:** Gernerally speaking, timesteps should neither be too small, ensuring data smoothly turn into normal distribution in the diffusion process, nor too large, which lead to slow sampling speed and unstable training. Experimentally, we set $T = 500$. Additionally, sampling method such as strided schedule (Nichol & Dhariwal, 2021) and DDIM (Song et al., 2020) can mitigate the cost of sampling and thus allow larger timesteps.

We set the forward process variances to constants increasing linearly from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$, matching previous work (Ho et al., 2020). Also, we find cosine schedule (Nichol & Dhariwal, 2021) does not work well in our case, resulting in low quality images.

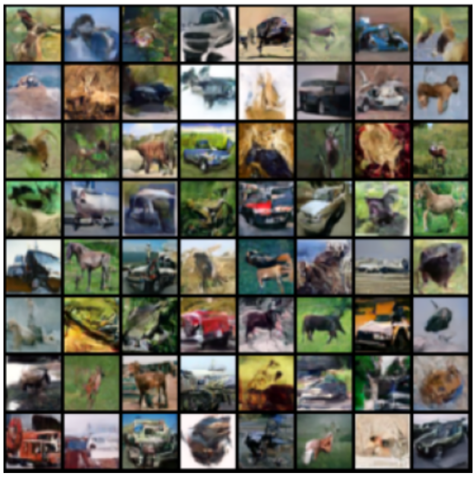

*Figure 3.* Generated samples on unconditional CIFAR10 (32*32)

**Neural Networks:** The task of neural networks $\epsilon_\theta\left(x_t, t\right)$ in the model is to take in noise image $x_t$ and time $t$ to predict the added noise $\epsilon$. Here, we employ U-net (Ronneberger et al., 2015) and integrate time embeddings, ResNet blocks, attention and group normalization in the network. Futhermore, the architecture of U-net can be modified (e.g. rescaling residual connections to $1/\sqrt{2}$) to enhance the performance of the model (Dhariwal & Nichol, 2021).

**Image Sizes:** 'CUDA out of memory' is a common problem when training diffusion models from scratch, especially when image sizes grow larger. For example, we fail to train 128*128 images with moderate batchsize on 3090Ti (24GB). In the experiment, we train on 32*32 images with batchsize of 128 and 64*64 images with batchsize of 64. One solution is CDM (Ho et al., 2021), which employs noise conditioning augmentation to scale up generation resolution and quality.

**Pre-training:** Pre-training is necessary before fine-tuning. Experimentally, we start with CIFAR10 (Krizhevsky et al., 2009), which captures characteristic of common objects and is relatively easy to learn from.

### 4.2. Results

Trained on CIFAR10 dataset, the diffusion model generates similar images (as shown in fig.3). It is worth noting that by altering the timesteps of sampling process, with other things unchanged, we aquire images of different contrast (as shown in fig.4).

Futhermore, we fine tune the diffusion model on oil painting dataset and anime face dataset, generating images of specific styles (as shown in fig.5 and fig.6).

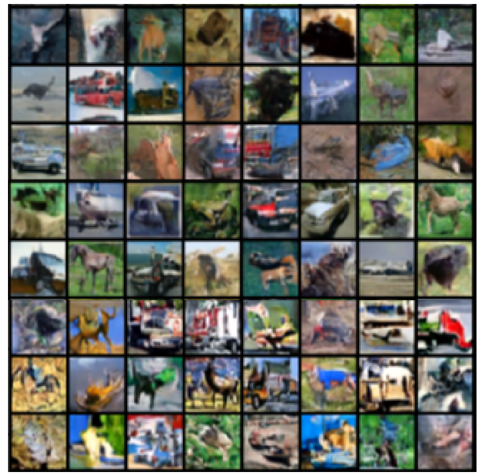

*Figure 4.* Generated samples with different sampling timesteps, the upper 3 rows $T = 300$, the middle 2 row $T = 500$, the lower 3 rows $T = 700$, all trained with $T = 500$.

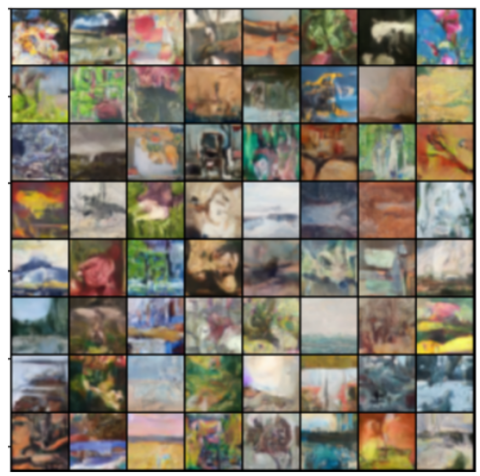

*Figure 5.* Generated samples on oil painiting dataset (32*32)

*Figure 6.* Generated samples on anime face dataset (64*64)

## 5. Conclusions

We provide a comprehensive look at diffusion models from various angles. We begin with developments, mechanism, applications and improvments of diffusion models. We also make comparisons between diffusion models and various generative models, demonstrating the edge of diffusion models. Finally, we fine tune diffusion models to generate pictures of specific style.

## References

Arjovsky, M. and Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1701.04862, January 2017. doi: 10.48550/arXiv.1701.04862.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein GAN. *arXiv e-prints*, art. arXiv:1701.07875, January 2017. doi: 10.48550/arXiv.1701.07875.

Austin, J., Johnson, D. D., Ho, J., Tarlow, D., and van den Berg, R. Structured denoising diffusion models in discrete state-spaces. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/958c530554f78bcd8e97125b70e6973d-Paper.pdf.

Avrahami, O., Lischinski, D., and Fried, O. Blended Diffusion for Text-driven Editing of Natural Images. *arXiv e-prints*, art. arXiv:2111.14818, November 2021. doi: 10.48550/arXiv.2111.14818.

Brock, A., Donahue, J., and Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv e-prints*, art. arXiv:1809.11096, September 2018. doi: 10.48550/arXiv.1809.11096.

Dhariwal, P. and Nichol, A. Diffusion Models Beat GANs on Image Synthesis. *arXiv e-prints*, art. arXiv:2105.05233, May 2021. doi: 10.48550/arXiv.2105.05233.

Dinh, L., Krueger, D., and Bengio, Y. NICE: Non-linear Independent Components Estimation. *arXiv e-prints*, art. arXiv:1410.8516, October 2014. doi: 10.48550/arXiv.1410.8516.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1406.2661, June 2014. doi: 10.48550/arXiv.1406.2661.

Ho, J. and Salimans, T. Classifier-Free Diffusion Guidance. *arXiv e-prints*, art. arXiv:2207.12598, July 2022. doi: 10.48550/arXiv.2207.12598.

Ho, J., Chen, X., Srinivas, A., Duan, Y., and Abbeel, P. Flow++: Improving Flow-Based Generative Models with Variational Dequantization and Architecture Design. *arXiv e-prints*, art. arXiv:1902.00275, February 2019. doi: 10.48550/arXiv.1902.00275.

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2006.11239, June 2020. doi: 10.48550/arXiv.2006.11239.

Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., and Salimans, T. Cascaded Diffusion Models for High Fidelity Image Generation. *arXiv e-prints*, art. arXiv:2106.15282, May 2021. doi: 10.48550/arXiv.2106.15282.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. Analyzing and Improving the Image Quality of StyleGAN. *arXiv e-prints*, art. arXiv:1912.04958, December 2019. doi: 10.48550/arXiv.1912.04958.

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. *arXiv e-prints*, art. arXiv:1312.6114, December 2013. doi: 10.48550/arXiv.1312.6114.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. RePaint: Inpainting using Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2201.09865, January 2022. doi: 10.48550/arXiv.2201.09865.

Nichol, A. and Dhariwal, P. Improved Denoising Diffusion Probabilistic Models. *arXiv e-prints*, art. arXiv:2102.09672, February 2021. doi: 10.48550/arXiv.2102.09672.

Razavi, A., van den Oord, A., and Vinyals, O. Generating Diverse High-Fidelity Images with VQ-VAE-2. *arXiv e-prints*, art. arXiv:1906.00446, June 2019. doi: 10.48550/arXiv.1906.00446.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models, 2021.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, art. arXiv:1505.04597, May 2015. doi: 10.48550/arXiv.1505.04597.

Rudolph, M., Wandt, B., and Rosenhahn, B. Same Same But DifferNet: Semi-Supervised Defect Detection with Normalizing Flows. *arXiv e-prints*, art. arXiv:2008.12577, August 2020. doi: 10.48550/arXiv.2008.12577.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Kamyar Seyed Ghasemipour, S., Karagol Ayan, B., Mahdavi, S. S., Gontijo Lopes, R., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *arXiv e-prints*, art. arXiv:2205.11487, May 2022. doi: 10.48550/arXiv.2205.11487.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2023. doi: 10.1109/TPAMI.2022.3204461.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved Techniques for Training GANs. *arXiv e-prints*, art. arXiv:1606.03498, June 2016. doi: 10.48550/arXiv.1606.03498.

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *arXiv e-prints*, art. arXiv:1703.05921, March 2017. doi: 10.48550/arXiv.1703.05921.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv e-prints*, art. arXiv:1503.03585, March 2015. doi: 10.48550/arXiv.1503.03585.

Song, J., Meng, C., and Ermon, S. Denoising Diffusion Implicit Models. *arXiv e-prints*, art. arXiv:2010.02502, October 2020. doi: 10.48550/arXiv.2010.02502.

Song, Y. and Ermon, S. Generative Modeling by Estimating Gradients of the Data Distribution. *arXiv e-prints*, art. arXiv:1907.05600, July 2019. doi: 10.48550/arXiv.1907.05600.

Weng, L. What are diffusion models? *lilianweng.github.io*, Jul 2021. URL https://lilianweng.github.io/posts/2021-07-11-diffusion-models/.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., and Yang, M.-H. Diffusion Models: A Comprehensive Survey of Methods and Applications. *arXiv e-prints*, art. arXiv:2209.00796, September 2022. doi: 10.48550/arXiv.2209.00796.