

Applications of Denoising Diffusion Probability Model in Image Processing

Renke He¹, Ruiqi Jiang^{2,4,*}, Zhilin Miao³

¹School of Sciences, Tianjin University of Commerce, Tianjin, China

²School of Mathematical Sciences, Nanjing Normal University, Jiangsu, China

³School of Mathematics, Hunan University, Hunan, China

⁴06230307@njnu.edu.cn

*corresponding author

Abstract. In 2020, Jonathan Ho et al. from the University of Berkeley proposed the Denoising Diffusion Probabilistic Models (DDPM), which improved on the shortcomings of the previous-generation Deformable Part Models (DPM). At the same time, they also surpassed previous generative models in image synthesis effects by using noise prediction, such as Generative Adversarial Networks (GAN), Variational Auto-Encoders (VAE), flow-based models and energy-based models, etc. After that, the denoising diffusion probabilistic models gradually received more discussions and research. In 2022, OpenAI launched the DALL-E 2, realizing Text-to-Image generation combined with the denoising diffusion probabilistic model. Two years later, with the release of Sora, the first Text-to-Video large models based on DALL-E, the denoising diffusion probabilistic models have received unprecedented attention in the multi-modal field. This article first provides the background and development process of the denoising diffusion probabilistic models. Secondly, it introduces the algorithmic principles of the diffusion models. Following that, it lists the applications of the models in the area of images. In the last section, it expounds on the advantages, disadvantages and future trends of the denoising diffusion probabilistic models.

Keywords: Generative Models, Diffusion Models, Noise Prediction, Text to Image, Multi-modality.

1. Introduction

With the rise of various types of AI since recent years, models for image generation have emerged. Some of the mainstream text-to-image models such as DALL-E 2, Stable Diffusion and Imagen adopt the diffusion model [1]. In paper [2], the authors point out that the core idea of the diffusion model originates from the diffusion phenomenon in physics, where molecules of a substance are transferred from a region of high concentration to a region of low concentration until a uniform distribution is achieved. This is a systematic and slow process of forward diffusion by destroys the structure in the data distribution. Then, the process of reverse diffusion is learned to achieve the restoration of the data in the structure. In the field of computer vision, the denoising diffusion probabilistic model is a generative model based on the diffusion process, where the forward process adds Gaussian noise to the original data, and the reverse process removes the noise to achieve image generation. This is undoubtedly a very

valuable and promising research direction for the hot field of AI generated content in recent years. The powerful capability demonstrated by the model has allowed it to quickly attract a large and extensive amount of research, and it has now become quite popular in related fields.

The development process of the denoising diffusion probabilistic model can be divided into the following three stages:

1. This classical model was first presented in the original paper [1]. In this paper, the authors presented the process of the diffusion model in detail based on the paper [2]. Since then, denoising diffusion probability models have only gradually received more attention.

2. The subsequent paper [3] affirmed the high-quality image generation capability of Denoising Diffusion Probabilistic Models (DDPM), but pointed out the inadequacy of its sampling speed. In order to accelerate the sampling, more efficient Denoising Diffusion Implicit Models (DDIM) are proposed in the paper, and the sampling speed is significantly improved. In DDPM, both forward additive and reverse denoising processes rely on Markov chains, resulting in slower sampling speed. DDIM, on the other hand, gets rid of this limitation and makes it possible to jump-step denoising during sampling, thus speeding up the generation process. Then paper [4] makes a simple modification to DDMP to achieve competitive log-likelihood while maintaining high image quality, and also shows that the sample quality and likelihood can be scaled with model capacity.

3. Once Google, OpenAI and others successfully trained the Big Model, the model hit the big time and with its unique advantages, it has achieved remarkable results in many fields. Now that the model is in its booming phase, it has been witnessed that DDPM can be combined with other generative models or machine learning models. For example, combining DDPM with generative adversarial networks (GAN) can improve the quality and diversity of generated images; combining DDPM with reinforcement learning can enable more complex tasks such as robot control and game intelligence; Combining DDPM with spatiotemporal prediction models [5] can fully utilize the temporal dynamic information in the data, thereby processing complex dynamic image data. Not only that, the training and sampling part of DDPM is still time-consuming, and researchers are exploring new training methods to reduce the time and computational cost.

In this article, firstly, the forward and inverse processes of denoising diffusion probabilistic models are introduced from the perspective of probabilistic models in the second part, and the pseudo-code for training and sampling is given after briefly describing the process of deriving the loss function in the paper [1]. Then in the third part the applications of denoising probabilistic diffusion model in image restoration, image generation, image enhancement, image denoising and biological structure design are listed in order with the literature. Then, in the fourth part, several advantages as well as disadvantages of the diffusion model are presented from the perspectives of data processing and training and sampling, respectively, and the section concludes with a vision for the future of diffusion models.

2. The Algorithm Principles of Denoising Diffusion Probabilistic Models

In this section, the probability models of the forward process and the reverse process in the diffusion models are introduced in detail. Also, the loss function derived by optimizing the KL-divergence of the probability distributions of the original image and the generated image is presented. Finally, the pseudo-codes for training and sampling of the diffusion models are given.

2.1. Forward Process

The forward process can be expressed as

$$x_0 \xrightarrow[\beta_1]{\text{add noise } \epsilon_1} \dots \xrightarrow[\beta_{t-1}]{\text{add noise } \epsilon_{t-1}} x_{t-1} \xrightarrow[\beta_t]{\text{add noise } \epsilon_t} x_t \xrightarrow[\beta_{t+1}]{\text{add noise } \epsilon_{t+1}} \dots \xrightarrow[\beta_T]{\text{add noise } \epsilon_T} x_T, \quad (1)$$

where the original image data $x_0 \sim q(x_0)$, noises $\epsilon_1, \dots, \epsilon_T \sim \mathcal{N}(0, I)$ and hyper-parameters $\beta_1, \dots, \beta_T \in (0, 1)$ satisfy

$$x_t = \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} \epsilon_t, \quad (2)$$

By re-parameterizing equation (2), the forward process kernel can be obtained

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I\right). \quad (3)$$

In the actual process, instead of gradually adding noise to x_{t-1} to obtain x_t , the probability distribution of x_t at any time is directly obtained from the given x_0 , i.e.

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{(1 - \bar{\alpha}_t)}\epsilon \Leftrightarrow q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I\right). \quad (4)$$

where $\alpha_t := 1 - \beta_t$, $\bar{\alpha}_t := \prod_{n=1}^t \alpha_n$, $\epsilon \sim \mathcal{N}(0, I)$.

2.2. Reverse Process

The forward process can be expressed as

$$x_T \xrightarrow[\epsilon_\theta(x_T, T)]{\text{denoise}} \cdots \xrightarrow[\epsilon_\theta(x_{t+1}, t+1)]{\text{denoise}} x_t \xrightarrow[\epsilon_\theta(x_t, t)]{\text{denoise}} x_{t-1} \xrightarrow[\epsilon_\theta(x_{t-1}, t-1)]{\text{denoise}} \cdots \xrightarrow[\epsilon_\theta(x_1, 1)]{\text{denoise}} x_0, \quad (5)$$

where $x_T \sim p(x_T) = \mathcal{N}(x_T; 0, I)$, θ is the model parameter from neural network training, and the label of the noise prediction model $\epsilon_\theta(x_t, t)$ is selected as a Gaussian distribution.

Define the reverse process kernel as

$$p_\theta(x_{t-1}|x_t) := \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right). \quad (6)$$

The label of the reverse process kernel $p_\theta(x_{t-1}|x_t)$ is $q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)$, and

$$\tilde{\beta}_t := \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t, \quad \tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0. \quad (7)$$

In paper [1], the authors fix the covariance matrix of the inverse process kernel as a parameter that does not require learning, i.e. let $\Sigma_\theta(x_t, t) = \sigma_t^2 I$, where $\sigma_t^2 = \tilde{\beta}_t := \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$. As to the mean of the reverse process kernel, its relation with the noise prediction models can be obtained through the optimization,

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{(1 - \bar{\alpha}_t)}} \epsilon_\theta(x_t, t) \right). \quad (8)$$

Moreover, the papers [3, 4] improve the setting of the covariance matrix $\Sigma_\theta(x_t, t)$ of the reverse process kernel on the basis of the paper [1], which enhances the training speed and sampling quality to a certain extent.

2.3. Optimization Process

The optimization objective is to make the probability distribution $p_\theta(x_0)$ of the generated image as close as possible to the original image $x_0 \sim q(x_0)$. This is equivalent to minimizing the KL-divergence $D_{KL}(q(x_0) || p_\theta(x_0))$ which is used to approximate $q(x_0)$ with $p_\theta(x_0)$.

In paper [1], the authors transform minimizing $D_{KL}(q(x_0) || p_\theta(x_0))$ into minimizing the KL-divergence between the reverse process kernel and its label. Finally, they introduce the noise predictor and simplify it to obtain the loss function as

$$L_{\text{simple}}(\theta) = \mathbb{E}_{t, q(x_0, \epsilon)} \left(\left\| \epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{(1 - \bar{\alpha}_t)}\epsilon, t \right) \right\|_2^2 \right), \quad (9)$$

where $t \sim \text{Uniform}(\{1, \dots, T\})$.

2.3.1. Training. This process is a loop. Firstly, extract an image $x_0 \sim q(x_0)$ from the database. Then, sample a number $t \sim \text{Uniform}(\{1, \dots, T\})$ and a noise $\epsilon \sim \mathcal{N}(0, I)$ respectively, where T is generally a relatively large integer, such as 1000. Subsequently, take gradient descent step on loss function, i.e.

$$\nabla_{\theta} \left\| \epsilon - \epsilon_{\theta} \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{(1 - \bar{\alpha}_t)} \epsilon, t \right) \right\|^2.$$

Finally, this loop stops when training converges.

2.3.2. Sampling. This process first samples a Gaussian distribution $x_T \sim \mathcal{N}(0, I)$, where T is the integer in training process. Next, take iterations from T to 0 on $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{(1-\bar{\alpha}_t)}} \epsilon_{\theta}(x_t, t)) + \sigma_t z$, where $\epsilon_{\theta}(x_t, t)$ is from training process and $z \sim \mathcal{N}(0, I)$ if $t > 1$, else $z = 0$. Ultimately, return x_0 as the generated graph.

3. Applications of Denoising Diffusion Probabilistic Models

3.1. Inpainting

In the paper [6], the authors proposed repaint, an image restoration repair method based on denoising diffusion probabilistic model. The method starts with complete noise and denoises the missing regions and fills the missing parts based on the known regions of the image. The method provides a better quality of repair in extreme cases compared to other artificial neural networks such as GAN. Not only that, the denoising diffusion based probabilistic model is capable of restoring more types of images and provides high diversity. The method has a greater potential and possibility in the field of art and heritage restoration, which is of great significance for the preservation of works.

3.2. Image Generation

In image generation, denoising diffusion probabilistic models have a wide range of applications, such as synthetic medical image generation systems based on denoising diffusion probabilistic models. In paper [7], it is stated that medical imaging analysis plays a major role in healthcare and generates a considerable amount of data that can be used to study complex diseases and transmission pathways. These data undoubtedly provide a large amount of training data for denoising diffusion probabilistic models, which are ultimately applied to the analysis of medical images, greatly improving the efficiency of diagnosis and the accuracy of assessment.

In addition to medical and other professional fields, there are more applications in daily life such as Stable Diffusion and other AI image generation products. The application through the algorithm iteration, the AI generation of picture fineness greatly improved, and can be completed in a very short period of time output. It can be used for the production of special effects in films and TV series, animation production and so on. It can quickly generate complex scenes, characters and actions, providing creators with more creativity and possibilities, and reducing production costs and time.

3.3. Image Enhancement

Since the underwater environment is very different from the atmospheric environment, the distortion problem of underwater images is very common, which leads to the quality of underwater images often being unsatisfactory. Inspired by DDPM, the authors of paper [8] proposed a DDPM-based underwater image enhancement method, UW-DDPM. UW-DDPM is trained on paired datasets, and two networks are used to complete the image denoising and the image distribution transformation, which effectively improves the quality of underwater images. Test results on real underwater images show that UW-DDPM achieves significant improvement over existing models in both visual effect and evaluation

metrics. It is foreseeable that this image enhancement method will undoubtedly help more underwater activities in the contemporary era when ocean exploration is becoming more and more frequent.

3.4. Image Denoising

The quality of images is of paramount importance, especially in the medical field, and the advent of PET has been a major revolution in medical imaging. In paper [9], the authors evaluated PET images for denoising and found that denoising methods under the DDPM-based framework produce better results than those based on other networks such as Generative Adversarial Networks GAN, and can be better combined with a priori information to achieve optimal performance and further reduce the uncertainty in the denoising process. Improvement in PET image quality can provide more valuable information, which can produce more accurate results, and this is critical for the early detection and diagnosis of major diseases. In addition, DDPM-based image denoising methods are expected to play an important role in many other fields, such as images taken in harsh environments, historical image restoration, etc.

3.5. Bioengineering

In the field of bioengineering, protein engineering has attracted wide attention. Designing protein structures capable of fulfilling specific functions is difficult due to the great differences in the 3D structures of various proteins and the arrangement of amino acids. According to the paper [10], denoising diffusion probabilistic models can generate the structure and sequence of proteins on a much larger molecular scale than previously modelled, compared to conventional methods. It helps to achieve the design of proteins that can fulfil the desired function and has an important role for bioengineering.

4. The future competitiveness of denoising diffusion models

4.1. Advantages of denoising diffusion models.

Sections should be numbered with a dot following the number and then separated by a single space:

4.1.1. High-quality data generation. In terms of image generation, the DDPM model can synthesize high-quality, realistic and diverse images.

4.1.2. Have excellent data denoising ability. It can effectively recover the original signal from noisy data.

4.1.3. Effective data augmentation. By generating new samples for existing datasets, it helps improve the generalization ability of machine learning models and reduce the excessive dependence of models on specific datasets. This is especially useful when training large neural networks, as they usually require a large amount of data to learn.

4.1.4. Powerful anomaly detection ability. Since DDPM can learn the normal distribution of data, they can effectively identify outliers that do not conform to this distribution.

Disadvantages of denoising diffusion models:

4.2. Disadvantages of denoising diffusion models

4.2.1. High computational cost: Denoising diffusion probabilistic models usually require a large amount of computing resources and time for training and sampling. During the training process, many steps of the Markov chain need to be simulated, and for each step, forward and backward propagation calculations of the model are required, which leads to a long training time.

4.2.2. Slow sampling speed: When generating samples, denoising diffusion probabilistic models also need multiple iterations to obtain high-quality samples. Compared with other generative models such as GANs, the sampling speed is slower.

4.2.3. Large memory usage: Due to the need to store intermediate results and model parameters, denoising diffusion probabilistic models may occupy a large amount of memory space during training and sampling.

4.2.4. Sensitive to hyperparameters: The performance of the model is relatively sensitive to the selection of hyperparameters such as learning rate and noise schedule. Inappropriate hyperparameter settings may lead to unstable model training, slow convergence speed, or poor generation quality. A large number of experiments and tunings are required to find a suitable combination of hyperparameters.

4.3. Future prospects of denoising diffusion models

4.3.1. Integration with other technologies. In paper [1], the authors propose that due to the seemingly excellent inductive bias of diffusion models on image data, they look forward to studying their applications in other data models and as components of other types of generative models and machine learning systems. This provides the possibility for the DDPM model to be applied in more fields and combined with other models. For example, combining DDPM with GAN can improve the quality and diversity of generated images. By combining other machine learning techniques such as reinforcement learning and self supervised learning, more powerful hybrid models can be created. It can also be combined with language models (such as GPT series) to generate more accurate and creative multimodal content based on text descriptions.

4.3.2. Multimodal data generation. Researchers believe that it is possible to explore how DDPM can be applied to the generation and processing of multimodal data, such as audio, video, etc. The generation of multimodal data will bring more possibilities to fields such as creative industries and entertainment.

4.3.3. Expand the scope of application. DDPM is expected to be applied in other fields, such as medical image analysis, industrial design, virtual reality, and augmented reality. Assist in disease diagnosis in the medical field, generate virtual medical images for training, and quickly generate conceptual designs in industrial design.

4.3.4. Further improvement in generation quality and realism. With the deepening of research and the advancement of technology, the DDPM model is expected to generate more realistic and detailed images or other data. By improving the model structure, training methods, and hyperparameter adjustments, the quality of generated results can be enhanced to better conform to human visual and cognitive habits.

4.3.5. Model compression and efficiency optimization. In order to run more efficiently in practical applications, future efforts may focus on model compression and optimization, reducing computational resource requirements and improving generation speed. This includes exploring more efficient network architectures, quantization methods, etc. to enable it to run on resource constrained devices.

4.3.6. More in-depth theoretical research. Further explore the theoretical basis of the DDPM model, such as deeper integration with mathematical theories such as stochastic differential equations, to promote the understanding and development of generative models.

5. Conclusion

This paper reviews the development history of denoising diffusion probabilistic models, analyzes some of the main application fields of this model, and expounds on the significant advantages and unique characteristics of this model at present. At present, due to its fundamental algorithmic advantages, the

denoising diffusion probabilistic model has high-quality generated data and extremely excellent denoising ability compared to other models, and its development momentum is very rapid. It is worth mentioning that this model can not only perform difficult tasks in professional fields, but also provides more possibilities in daily life scenarios. It can be predicted that in the era of big data, denoising diffusion probabilistic models will play a more important role in modern technology. This will be a further exploration of its potential and also a challenge to itself. Accelerating the integration with other technologies and fields and improving its own sampling efficiency will better exert the capabilities of this model.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, Vol. 33. 6840–6851.
- [2] Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*. 2256–2265.
- [3] Song, J., Meng, C., & Ermon, S. (2020). Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*.
- [4] Nichol, A., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*. 8162–8171.
- [5] Cachay, S. R., Zhao, B., Joren, H., & Yu, R. (2022). DYffusion: A Dynamics-informed Diffusion Model for Spatiotemporal Forecasting. In *Neural Information Processing Systems*.
- [6] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Gool, L. V. (2022). RePaint: Inpainting using Denoising Diffusion Probabilistic Models. In *Conference on Computer Vision and Pattern Recognition*.
- [7] Khader, F., Mueller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarbuerger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baessler, B., Foersch, S., Stegmaier, J., Kuhl, C., Nebelung, S., Kather, J. N., & Truhn, D. (2023). Denoising diffusion probabilistic models for 3D medical image generation. In *Scientific Reports*.
- [8] Lu, S., Guan, F., Zhang, H., & Lai, H. (2023). Underwater image enhancement method based on denoising diffusion probabilistic model. In *Social Science Research Network*.
- [9] Gong, K., Johnson, K. A., Fakhri, G. E., Li, Q., & Pan, T. (2022). PET image denoising based on denoising diffusion probabilistic models. In *European Journal of Nuclear Medicine and Molecular Imaging*.
- [10] Anand, N., & Achim, T. (2022). Protein structure and Sequence Generation with Equivariant Denoising Probabilistic Diffusion Models. In *Neural Information Processing Systems*.