Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

DENOISING DIFFUSION PROBABILISTIC MODEL WITH WAVELET PACKET TRANSFORM FOR FINGERPRINT GENERATION

Li Chen and Huah Yong Chan

(Received: 29-Jun.-2024, Revised: 9-Aug.-2024, Accepted: 12-Aug.-2024)

ABSTRACT

The majority of contemporary fingerprint synthesis is based on the Generative Adversarial Network (GAN). Recently, the Denoising Diffusion Probabilistic Model (DDPM) has been demonstrated to be more effective than GAN in numerous scenarios, particularly in terms of diversity and fidelity. This research develops a model based on the enhanced DDPM for fingerprint generation. Specifically, the image is decomposed into sub-images of varying frequency sub-bands through the use of a wavelet packet transform (WPT). This method enables DDPM to operate at a more local and detailed level, thereby accurately obtaining the characteristics of the data. Furthermore, a polynomial noise schedule has been designed to replace the linear noise strategy, which can result in a smoother noise-addition process. Experiments based on multiple metrics on the datasets SOCOFing and NIST4 demonstrate that the proposed model is superior to existing models.

KEYWORDS

Diffusion model, Fingerprint, Image processing, Wavelet packet transform.

1. INTRODUCTION

The singularity and immutability of fingerprints render them an indispensable biometric trait of humans, which is extensively utilized in a multitude of contexts including identity authentication, criminal investigation, medical research and so forth [1]–[3]. Nevertheless, the establishment of such a fingerprint system necessitates a substantial quantity of fingerprint samples and direct fingerprint collection is constrained by equipment, environmental or legal considerations. In response to this challenge, the advent of computer vision has led to the emergence of synthetic fingerprint technology. However, the intricacies of fingerprints are considerable, encompassing a multitude of diverse pattern types, including whorls, loops and arches [4].

Most early model-based methods typically synthesize or reconstruct fingerprints based on minutiae templates or shallow model. While these methods have been demonstrated to be effective in synthesizing fingerprint structure and texture [5]–[8], they exhibit low automation levels and poor model adaptability and scalability. Furthermore, deep learning-based generative models have gradually become the dominant models for fingerprint generation, employing more complex network structures, such as auto-encoders and GANs [9]–[13]. However, images generated by VAEs are often quite blurry and the shortcomings of GAN models, such as mode collapse and difficulty in training, cannot be ignored either.

The recent emergence of Denoising Diffusion Probabilistic Models (DDPMs) [14] has ushered in a novel technical approach to image synthesis. This model is based on the Markov chain, which gradually introduces noise to the data in the forward stage until the data is corrupted and becomes completely Gaussian noise. In the reverse stage, the Gaussian noise is then restored to the original data. The model must ensure that the reverse Markov chain closely resembles the forward process during optimization. It has demonstrated remarkable efficacy in the generation of images, videos and other forms of data, particularly in the domain of image synthesis, as evidenced by its performance in DALL-E 2 [15] and Stable Diffusion [16].

This study proposes a novel fingerprint-generation method based on an enhanced model of DDPM. Specifically, the original image is decomposed into sub-images of varying frequency bands through wavelet packet transform. These sub-images contain information of different scales and directions, which enables DDPM to operate at a more local and detailed level, thereby improving the stability of

L. Chen and H.Y.Chan are with School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia. Emails: chenli0926@student.usm.my and hychan@usm.my

training and the quality of generated images. Moreover, in consideration of the potential for the linear noise schedule in the original DDPM to result in the corruption of data too quickly in the forward phase, this study devised a polynomial noise schedule to facilitate a more gradual and less disruptive noise-addition process, which prevents an abrupt change in noise level. Experimental results have demonstrated the effectiveness of this approach in both SOCOFig [17] and NIST Special Database 4 [18].

2. RELATED WORK

Inspired by non-equilibrium statistical physics, [19] designed a parameterized Markov-chain model that slowly destroys data in the forward phase and gradually recovers data in the backward phase which is highly flexible and tractable. DDPM model further extended this idea to high quality tasks. The forward process commences with the real distribution of data $q(x_0)$ and incorporates a minimal quantity of Gaussian noise at each stage, gradually transforming the data into pure noise. The whole process can be represented by $q(x_t|(x_{t-1})$. In the generation process, DDPM starts with a standard Gaussian noise sample x_t, gradually removes the noise through the reverse process and finally generates a sample x_0, which can be defined as $p_\theta((x_{t-1}|(x_t)$.

2.1 The Principle of DDPM

In the forward process, let $x_0, x_1 \ldots\ldots x_n$ represent the gradual addition of Gaussian noise to an image until the image becomes completely noisy. This process can be described as shown in Figure 1.

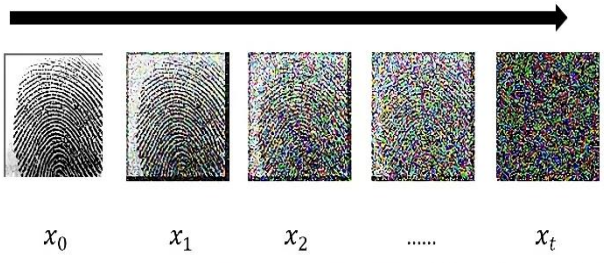


Figure 1. The forward process of DDPM.

It should be noted that the degree of noise added in each step is not uniform. It is controlled by weight β and its value must be gradually increased. For example, it normally ranges from 0.0001 to 0.02. The actual noise added is represented by α, which is defined as $\alpha_t = 1 - \beta_t$ to represent the weight of the noise at step t. Since the image obtained at each step depends on the image from the previous step plus the noise $z \sim N(0,1)$ of the current step, that is the model relies on a Markov chain, the image at any step can be obtained by $x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}z_t$. As the diffusion process continues, the noise content of the image will increase until it reaches a point where the entire image is comprised of noise. Due to the recursive nature of the model, the diffusion process does not occur in discrete steps during training. Instead, the model directly calculates the relationship between the original image and the noise image at any given step. This relationship is represented by the cumulative product of all steps, denoted by $\overline{\alpha_t}$. In the reverse process, the model must restore the noise to the image (Figure 2), Although it is challenging to determine $p_\theta(x_{t-1}|x_t)$ directly, x_0 and x_t are known in the forward process, which is available to calculate posterior probability $q(x_{t-1}|x_t, x_0)$ to estimate the mean and variance of reverse-process distribution. After Bayesian transformation, the goal can be transformed into calculating the distribution (2), (3), (4).

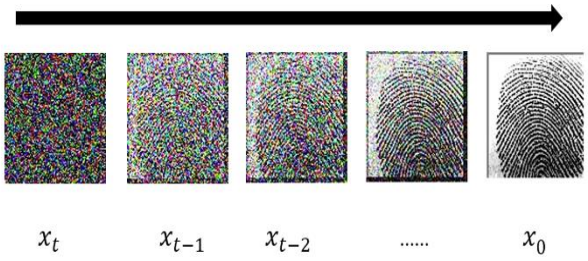


Figure 2. The reverse process of DDPM.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

$$q(x_{t-1}|x_t, x_0) = q(x_t|x_{t-1}, x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \quad (1)$$

$$q(x_{t-1}|x_0) = \sqrt{\overline{\alpha_{t-1}}}x_0 + \sqrt{1-\overline{\alpha_{t-1}}}z_{t-1} \sim N\left(\sqrt{\overline{\alpha_{t-1}}}x_0, 1-\overline{\alpha_{t-1}}\right) \quad (2)$$

$$q(x_t|x_0) = \sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}z_t \sim N\left(\sqrt{\overline{\alpha_t}}x_0, 1-\overline{\alpha_t}\right) \quad (3)$$

$$q(x_t|x_{t-1}, x_0) = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}z_t \sim N\left(\sqrt{\alpha_t}x_{t-1}, 1-\alpha_t\right) \quad (4)$$

Since all three distributions are normal distributions, but with different means and variances, according to the normal distribution probability density function, the mean (μ) and variance (σ) of the reverse-process distribution can be obtained:

$$\sigma^2 = \frac{\alpha_t}{\beta_t} + \frac{1}{1-\overline{\alpha_{t-1}}}{x_{t-1}}^2 \quad (5)$$

$$\mu = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{1-\overline{\alpha_t}}\epsilon_\theta\right) \quad (6)$$

Where ϵ_θ is the predicted value of z. Finally, the loss function can be defined as follows:

$$L_{simple}(\theta) = E_{t,x_0,z} \parallel \mathrm{z} - \epsilon_\theta\left(\sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}z, t\right) \parallel^2 \quad (7)$$

2.2 The Improvements of DDPM

Although DDPM has made considerable achievements, it still has shortcomings. [20] found that adding linear noise may cause the $\overline{\alpha_t}$ to approach 0 too quickly, which may damage data information quickly during training. It proposes adding a certain amount of cosine noise in the forward process, which can obtain better log-likelihoods. [21] established a noise signal in a learnable format and incorporated Fourier features into the input of the network, enabling the prediction of noise. These enhancements exceed the performance of autoregressive models. [22] put forth a methodology for dynamically adjusting the noise parameters with the objective of enhancing the denoising capabilities of the model, thereby improving the quality of the synthesized results. [23] introduced an Adversarial Purification approach, which incorporates noise and adversarial images into the forward process. In the reverse process, it is necessary to remove both the noise and the adversarial perturbation, while retaining the main content of the input image. This has the effect of improving the classification task. [24] presented a non-Markovian mechanism that renders the sampling process deterministic in the reverse process. This improvement has led to an improvement in the efficiency of the sampling process.

Some researchers have employed a combination of signal-processing techniques with DDPM. The wavelet transform is a widely utilized signal-processing technique that enables the decomposition of signals into components with varying frequencies and time resolutions. In the field of image processing, the wavelet transform is a valuable tool that can assist in the analysis and processing of information at different scales and frequencies within images. Its applications include tasks, such as image denoising, compression and feature extraction. In a recent study, [25] developed a model that combined wavelet and DDPM for 3D medical scans. This model performed a wavelet transform on the input to predict both wavelet coefficients and noise. [26] posited that the use of wavelet transform can mitigate the long-term inference issue by transferring the image reconstruction task from the spatial domain to the wavelet domain. [27] also used a similar concept, proposing a Stage-by-stage Wavelet Optimization Refinement Diffusion model, which takes wavelet transformation to improve the robustness of the model. The authors further argued that wavelet transform can facilitate the disentanglement of image content and features at varying scales, thereby enhancing the stability of the model-training process. [28] extracted the high-frequency and low-frequency information of the image after wavelet transformation in order to accelerate the training of the model without compromising the quality of the output.

3. DDPM WITH WAVELET PACKET TRANSFORM

This study implemented the following improvements based on DDPM. Firstly, the Wavelet Packet Transform (WPT) was initially introduced in the training stage. In this stage, the original data was transformed into wavelet packets to obtain sub-bands at different scales and frequencies. The diffusion

process was applied to the sub-bands of each scale and frequency to gradually add noise, which helps the model better extract data features and improve training stability. In the sampling stage, the original image was reconstructed using the inverse wavelet packet transform (IWPT). Moreover, a polynomial noise schedule was devised with the objective of reducing abrupt change of noise level. This strategy allows for the addition of noise in a more gradual and smoother manner, thereby reducing the potential for information loss.

3.1 Principle of Wavelet Packet Transform

The wavelet transform is a mathematical tool that employs various wavelets to decompose a signal into low-frequency and high-frequency components including the Morlet, Daubechies and Haar wavelets. Wavelet transform can provide both time and frequency information. It has a wide range of applications in image denoising, image compression, feature extraction and so forth. Wavelet packet transform is a generalization of wavelet transform, allowing further decomposition of the approximate and detailed parts of the signal. It provides finer frequency resolution by recursively decomposing all frequency bands of the signal. Specifically, for the discrete signal $f(t)$, its decomposition by wavelet packet transform can be expressed as follows:

$$W_{j,k}(t) = \sum_{t} f(t)\psi_{j,k}(t) \tag{8}$$

In this context, $\psi_{j,k}(t)$ represents the function of the wavelet packet transform, with j and k denoting the scale and position parameters, respectively. The reconstruction of the inverse wavelet packet transform can be defined as follows:

$$f(t) = \sum_{j} \sum_{k} W_{j,k}\psi_{j,k}(t) \tag{9}$$

The characteristics of WPT are employed to perform multi-scale processing on the model. Multi-scale analysis facilitates the capture of structural information within the data, thereby enhancing the quality of the denoising and reconstruction process. The addition of noise to different frequency components independently serves to render the distribution of noise more reasonable, which in turn improves the quality of the generated data.

3.2 Wavelet Packet Transform on DDPM

In the forward process, let $x \in \mathbb{R}^{D\times H\times W}$ denote the input data. During the experiment, the three-level wavelet packet transform of Haar was employed on the input data, with l and h indicating low-frequency and high-frequency sub-bands, respectively. Following the application of the wavelet packet transform to the first layer, four sub-images can be derived, which are: x_{ll}^1, x_{lh}^1, x_{hl}^1, x_{hh}^1, all of which are elements of $\mathbb{R}^{\frac{D}{2}\times\frac{H}{2}\times\frac{W}{2}}$. Subsequently, the two dimensions of x_{ll}^1 and x_{lh}^1 are concatenated in the width direction and the two dimensions of x_{hl}^1 and x_{hh}^1 are concatenated in the height direction. Finally, the two aforementioned structures are concatenated to obtain the matrix $x^1 \in \mathbb{R}^{D\times H\times W}$. As with a binary tree, it can be demonstrated that 16 sub-images can be obtained in the second layer and 64 sub-images in the third layer. Similar operations are then performed on the second and third layers to obtain x^2 and x^3, which are subsequently spliced from each level to finally yield $y \in \mathbb{R}^{3(D\times H\times W)}$, which is then used for DDPM processing. The complete training and sampling process of the model is as follows:

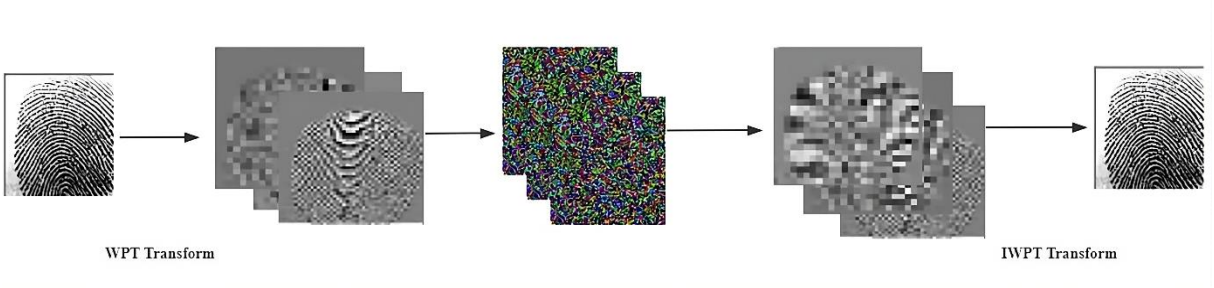


Figure 3. The pipeline of WPT diffusion.

Algorithm 1 Training

Repeat

$y_0 = WPT(x_0)$

T~Uniform (1,…,T)

$\epsilon \sim N(0,1)$

$q(y_t|y_{t-1}) = N\left(y_t, \sqrt{\alpha_t} y_{t-1}, (1-\alpha_t)I\right)$

Take gradient descent step on $\nabla_\theta \parallel \epsilon - \epsilon_\theta(y_t, t) \parallel^2$

Until converge

Algorithm 2 Sampling

$y_t \sim N(0, I)$

for t = T,…,1 do

z $\epsilon \sim N(0, I)$

$p_\theta = \left(y_{t-1} \middle| y_t, z, \epsilon_\theta(y_t, t)\right)$

$x_0 = IWPT\ (y_0)$

return x_0

During the training process, the input data is transformed by WPT into several sub-bands and all sub-band signals are spliced and trained in the WPT domain rather than in the original pixel domain. In the sampling process, the distribution of y_t is first determined to be $N(0, I)$ and the entire reverse process is then applied to obtain all sub-images. The image y_0 is finally reconstructed by performing an inverse wavelet packet transform (IWPT).

3.3 Polynomial Noise Schedule

Linear noise schedules may bring abrupt changes in noise level during training, which causes information disruption too quickly during training, especially for images with lower resolution (less than or equal to 64× 64) [20]. To avoid this phenomenon, this research designs a polynomial noise schedule, where this noise strategy is smoother in the process of adding, especially at $t = 0$ and $t = T$. It is milder in noise changes than the linear schedule and cosine schedule, which helps the stability of training and avoids instability of training caused by sudden noise changes.

$$\beta_t = \beta_0 + (\beta_T - \beta_0)(\frac{t}{T})^p \tag{10}$$

As Equation (10) shows, let β be the parameter to control the noise level, where β ranges form β_0 to β_T, t and T denote the current step and max step. By adjusting the power parameter p, the growth rate of β can be controlled, thereby affecting the smoothness of the cumulative product of $\overline{\alpha_t}$. p is the hyperparameter to adjust the smoothness of function. $p = 2$ is default setting during training, which is experimentally workable. Figure 4 shows that this noise schedule is smoother than the linear strategy and cosine strategy at the beginning and end of training.

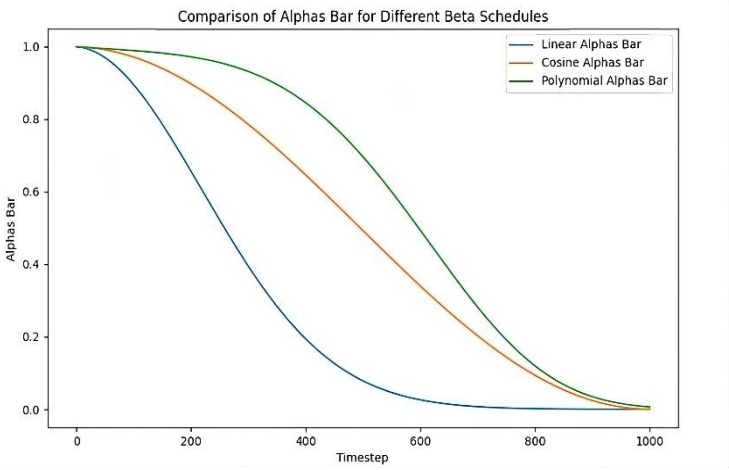


Figure 4. Compassion between different noise schedules.

3.4 Theoretical Analysis

In comparison to the conventional DDPM model, the methodology presented in this study offers enhanced capabilities in several key areas. Initially, the wavelet packet transform is capable of decomposing an image into sub-images of varying frequencies and directions, thereby encompassing the multi-scale characteristics of the original image. Moreover, it enables the separation of high-frequency and low-frequency components within an image, thereby facilitating the processing of these components in isolation and the reduction of blurring effects in the generated image. During the diffusion process, the model is capable of processing sub-images of varying frequencies in a manner that preserves and restores details and coarse information in a more optimal manner. In addition, the wavelet packet transform enables the separation of different image features, thereby enhancing the model's resilience to noise and other disturbances. Sub-images of varying frequency bands can be processed independently during the diffusion process, reducing mutual interference between different frequency bands. This improves the stability and quality of the generated image. Furthermore, polynomial noise schedule ensures a smoother way to add noise to each sub-image, which will contribute to the stability of model training.

4. EXPERIMENTS

The proposed model is evaluated with 200-epoch training on two distinct datasets: SOCOFing and NIST4 which are mentioned before. During the experiment, the architecture is based on U-Net included two attention blocks with a learning rate of 0.0001, β ranging from 0.0004 to 0.02 and a maximum step T of 1000. Finally, images with a resolution of 64×64 are generated.

4.1 Datasets

The Sokoto Coventry Fingerprint Dataset (SOCOFing) is a biometric fingerprint database created for academic-research purposes. This dataset comprises 6,000 fingerprint images with a resolution of 96×103 pixels, captured from 600 subjects of African descent. Each image is accompanied by a set of attributes, including gender, hand and finger name labels, as well as a sub-set of altered fingerprint images. This experiment employs all 6,000 real fingerprint images as training data.

NIST Special Database 4 (NIST 4) is a tool for evaluating fingerprint systems and contains 2000 8-bit grayscale fingerprint pairs, totaling 4000 images with 512×512 resolution. The database may be employed for the purposes of algorithm development, system training and testing. The experiment utilizes all 4,000 images as training data.

4.2 Progressive Generation

First, to verify the effectiveness of the proposed model, a sample was taken every 100 steps at T=1000 and 10 images were generated to observe the reverse diffusion process. As shown in Figure 5, which illustrates the reverse process of the proposed model, the image starts with noise, then evolves to the fingerprint outline and finally to a recognizable fingerprint texture. This result is indicative of the effectiveness of the model proposed in this research.

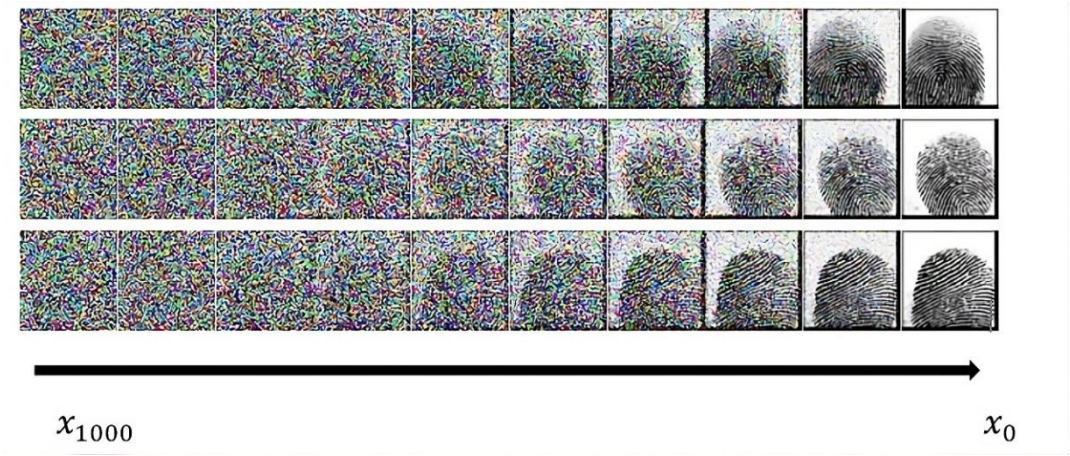


Figure 5. The reverse process of the proposed model.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

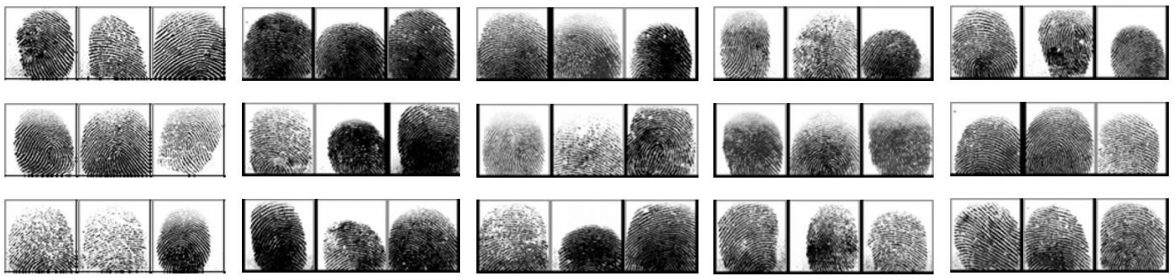
4.3 Results of Quality Evaluation

In the quality assessment, this sub-section firstly employs the Frechet Inception Distance (FID) [29] and Inception Score (IS) [30] as evaluation metrics on both SOCOFing and NIST4. FID and IS are two of the most commonly used metrics for evaluating the quality of images generated by generative models. Both require the utilization of pre-trained Inception v3 networks. The following models are employed for comparative purposes: DDPM (Vanilla) [14], DDPM (Wavelet) [25] and FingerGAN [12]. The results are presented in Table 1.

Table 1. The comparison between different models on IS and FID.

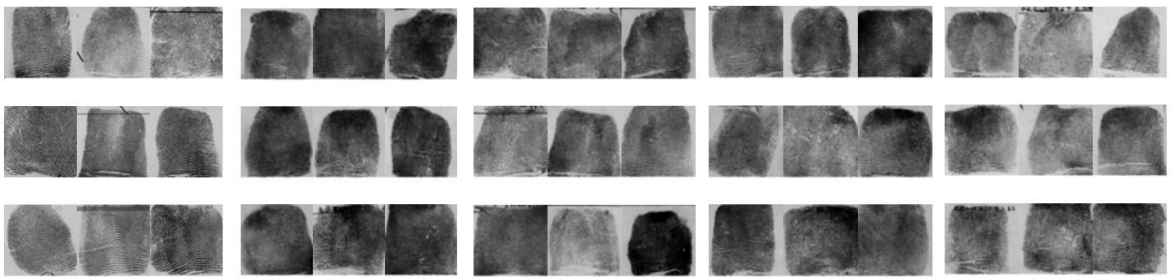
Model	SOCOFing		NIST4	
	FID	IS	FID	IS
DDPM(Vanilla) [14]	64.61	2.26 $\pm$0.16	62.71	2.69 $\pm$0.19
DDPM (Wavelet) [25]	62.96	2.33 $\pm$0.11	60.53	**2.73 $\pm$0.21**
FingerGAN [12]	70.35	2.12 $\pm$0.14	69.21	2.33 $\pm$0.12
Proposed Model	**61.64**	**2.42 $\pm$0.13**	**59.65**	2.72 $\pm$0.11

Table 1 demonstrates that the proposed model outperforms the other three models in terms of both FID and IS in the majority of scenarios. Nevertheless, the IS of DDPM (Wavlet) is marginally superior to the proposed model on the NIST4 dataset. This indicates that the proposed model is capable of generating high-quality and diverse results. Furthermore, this sub-section presents a comparative analysis of the images generated by all models, as illustrated in Figures 6 and 7. The data represented by the labels 1, 2, 3, 4, 5 in Figures 6 and 7 is real data, data generated by FingerGAN, data generated by DDPM, data generated by DDPM (wavelet) and data generated by the proposed model, respectively. In terms of visual quality, the four models demonstrate satisfactory performance on SOCOFing. However, the results produced by Finger GAN are not sufficiently diverse and the model proposed in this paper is more effective at generating fingerprint texture and structure. Furthermore, given the high resolution of the NIST4 original data, reaching at 512×512 resolution, it is not evident that there is a discernible visual difference between the four models at the resolution of 64×64.



(1) (2) (3) (4) (5)

Figure 6. The generated results from each model on SOCOFing.



(1) (2) (3) (4) (5)

Figure 7. The generated results from each model on NIST4.

In addition, as both FID and IS depend on the pretrained model, this sub-section also employs SSIM and MS-SSIM to assess the performance of the model. SSIM (Structural Similarity Index) is an index that gauges the similarity of two images. It considers the image's brightness, contrast and structure. MS-SSIM (Multi-scale SSIM) is an extension of SSIM that assesses the structural similarity of images at different scales. For the SOCOFing dataset, the generated original images were resized to 64×64 24-bit width. For the NIST4 data set, the generated original images were resized to 64×64 8-bit width. Finally,

the Average SSIM and Average MS-SSIM were calculated, as shown in Table 2.

As can be seen from Table 2, in general, the value of MS-SSIM is lower than that of SSIM. This discrepancy may be attributed to the fact that MS-SSIM considers the structural similarity of images at multiple scales, thereby providing a more comprehensive image-quality assessment. It can be observed that the proposed model exhibits better performance in both SSIM and MS-SSIM, particularly in SSIM on NIST4, where its score is considerably higher than those of other models. This may indicate that the proposed model is more adept at maintaining image quality. In addition, The MS-SSIM metric also demonstrates that the proposed model exhibits certain advantages when dealing with scenes involving image scaling or multi-scale characteristics.

Table 2. The comparison between different models on SSIM and MS-SSIM.

Model	SOCOFing		NIST4	
	SSIM	MS- SSIM	SSIM	MS- SSIM
DDPM(Vanilla) [14]	0.239	0.215	0.441	0.281
DDPM (Wavelet) [25]	**0.261**	0.227	0.452	0.262
FingerGAN [12]	0.227	0.203	0.346	0.272
Proposed Model	0.253	**0.239**	**0.495**	**0.295**

4.4 Ablation Study

In order to ascertain the effect of each modification to the model on its overall performance, an ablation study is conducted in this sub-section. The present study proposes the utilization of the Wavelet Packet Transformation (WPT) technique to decompose the original data into a number of sub-bands. Prior research has demonstrated that the application of wavelet transformation has resulted in a more stable training process. Wavelet packet transformation (WPT) is an extension of wavelet transformation (WT) offering a more detailed frequency decomposition. In comparison to WT, WPT further decomposes all components (including low-frequency and high-frequency) at each decomposition level, thus facilitating the extraction of more accurate features, which in turn improves the stability and quality of image generation. By training at 200 epochs, the loss changes between vanilla DDPM and the DDPM with WPT are shown in Figure 8. Obviously, these two models converge at approximately 20 epochs, but the loss of vanilla DDPM fluctuates. In contrast, the loss of DDPM with WPT is more stable.

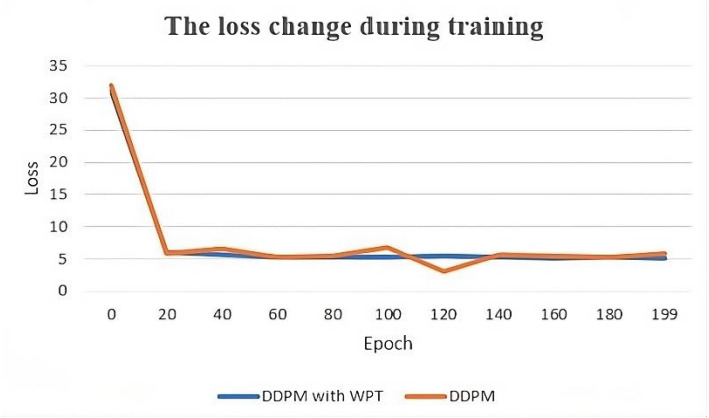


Figure 8. Loss comparison between DDPM and DDPM with WPT.

According to Table 3, the proposed improvements have led to an enhancement in the model performance, with WPT exhibiting the most notable improvement. Although the contribution of Polynomial noise schedule is not as great as that of WPT, it is demonstrably superior to the original linear schedule.

Table 3. The ablation study of each component.

Model	SOCOFing		NIST4	
	FID	IS	FID	IS
DDPM (Linear)	64.61	2.26 $\pm$0.16	62.71	2.69 $\pm$0.19
DDPM (Linear +WPT)	61.77	2.37 $\pm$0.21	60.22	2.71 $\pm$0.13
DDPM (Polynomial)	63.89	2.29 $\pm$0.14	61.98	2.70 $\pm$0.14
DDPM (Polynomial +WPT)	**61.64**	**2.42 $\pm$0.13**	**59.65**	**2.72 $\pm$0.11**

5. CONCLUSION AND FUTURE RESEARCH

This study proposes an improved version of DDPM for the task of fingerprint generation. Firstly, wavelet packet transform is used during training, which allows the model to better extract data features on different sub-bands. Secondly, a better noise schedule is developed to make noise addition smoother during the training process. The results of the experiments demonstrate that the proposed model exhibits superior performance compared to previous models. Nevertheless, the inference time remains comparable to that of the original DDPM, which represents a potential avenue for future research. Besides, the current equipment limits the resolution of the images generated by this model to 64×64, so the objective is to apply this model to a higher resolution in the future. It would be beneficial to explore the potential of this model for generating text and video as well.

REFERENCES

[1] J. Priesnitz et al., "An Overview of Touchless 2D Fingerprint Recognition," EURASIP Journal on Image and Video Processing, vol. 2021, no. 1, p. 8, Dec. 2021.

[2] J. Kolberg et al., "COLFISPOOF: A New Database for Contactless Fingerprint Presentation Attack Detection Research," Proc. of the 2023 IEEE/CVF Winter Conf. on Applications of Computer Vision Workshops (WACVW), pp. 653–661, Waikoloa, HI, USA, 2023.

[3] E. Prabakaran and K. Pillay, "Nanomaterials for Latent Fingerprint Detection: A Review," Journal of Materials Research and Technology, vol. 12, pp. 1856–1885, May 2021.

[4] M. Kücken M and A. Newell, "Fingerprint Formation," Journal of Theoretical Biology, vol. 235, no. 1, pp. 71–83, Jul. 2005.

[5] R. Cappelli et al., "Synthetic Fingerprint-image Generation," Proc. of the 15th Int. Conf. on Pattern Recognition (ICPR-2000), vol. 3, pp. 471–474, Barcelona, Spain, 2000.

[6] J. L. Araque et al., "Synthesis of Fingerprint Images," Proc. of the 2002 Int. Conf. on Pattern Recognition, vol. 2, pp. 422–425, DOI: 10.1109/ICPR.2002.1048329, Quebec City, QC, Canada, 2002.

[7] J. Feng and A. K. Jain, "Fingerprint Reconstruction: from Minutiae to Phase," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 209–223, 2011.

[8] T. Uz et al., "Minutiae-based Template Synthesis and Matching for Fingerprint Authentication," Computer Vision and Image Understanding, vol. 113, no. 9, pp. 979–992, Sep. 2009.

[9] M. Attia, "Fingerprint Synthesis *via* Latent Space Representation," Proc. of the 2019 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), pp. 1855–1861, Bari, Italy, 2019.

[10] O. Striuk and Y. Kondratenko, "Adaptive Deep Convolutional GAN for Fingerprint Sample Synthesis," Proc. of the 2021 IEEE 4th Int. Conf. on Advanced Information and Communication Technologies (AICT), pp. 193–196, DOI: 10.1109/AICT52120.2021.9628978, Lviv, Ukraine, 2021.

[11] P. Schuch et al., "De-convolutional Auto-encoder for Enhancement of Fingerprint Samples," Proc. of the 2016 6th Int. Conf. on Image Processing Theory, Tools and Applications (IPTA), pp. 1–7, Oulu, Finland, DOI: 10.1109/IPTA.2016.7821036, 2016.

[12] S. Minaee and A. Abdolrashidi, "Finger-GAN: Generating Realistic Fingerprint Images Using Connectivity Imposed GAN," arXiv, [Online], Available http://arxiv.org/abs/1812.10482, Dec. 2018.

[13] S. Seidlitz et al., "Generation of Privacy-friendly Datasets of Latent Fingerprint Images Using Generative Adversarial Networks," Proc. of the 16th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 345–352, DOI: 10.5220/0010251603450352, 2021.

[14] J. Ho et al., "Denoising Diffusion Probabilistic Models," Proc. of the 34th Int. Conf. on Neural Information Processing Systems, pp.6840-6851, Vancouver, BC, Canada, 2020.

[15] A. Ramesh et al., "Hierarchical Text-conditional Image Generation with CLIP Latents," arXiv, [Online], Available http://arxiv.org/abs/2204.06125, Apr. 2022.

[16] R. Rombach et al., "High-resolution Image Synthesis with Latent Diffusion Models," Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10674–10685, New Orleans, LA, USA, 2022.

[17] Y. I. Shehu et al., "Sokoto Coventry Fingerprint Dataset," arXiv, [Online], Available https://arxiv.org/abs/1807.10609, 2018.

[18] C. Watson, NIST Special Database 4, NIST 8-Bit Gray Scale Images of Fingerprint Image Groups, World Wide Web-Internet and Web Information Systems, [Online], Available: https://www.nist.gov/publications/nist-special-database-4-nist-8-bit-gray-scale-images-fingerprint-image-groups, 2008.

[19] J. Sohl-Dickstein et al., "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics," Proc. of the 32nd Int. Conf. on Machine Learning, vol. 37, pp.2256 - 2265, 2015.

[20] A. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," Proc. of the 38th Int. Conf. on Machine Learning, vol 139, pp.8162-8171, 2021.

[21] D. P. Kingma et al., "Variational Diffusion Models," Proc. of the 35th Int. Conf. on Neural Information

Processing Systems, pp. 21696-21707, DOI: 10.5555/3540261.3541921, 2021.
[22] R. San-Roman et al., "Noise Estimation for Generative Diffusion Models," arXiv, [Online], Available http://arxiv.org/abs/2104.02600, Sep. 2021.
[23] J. Wang et al., "Guided Diffusion Model for Adversarial Purification," arXiv, [Online], Available http://arxiv.org/abs/2205.14969, Jun. 2022.
[24] J. Song et al., "Denoising Diffusion Implicit Models," Proc. of the 2021 Int. Conf. on Learning Representations, Vienna, Austria, [Online], Available https://openreview.net/forum?id=St1giarCHLP, 2021.
[25] P. Friedrich et al., "WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis," arXiv, [Online], Available http://arxiv.org/abs/2402.19043, Feb. 2024.
[26] Y. Huang et al., "WaveDM: Wavelet-based Diffusion Models for Image Restoration," IEEE Transactions on Multimedia, vol. 26, pp. 7058-7073, DOI: 10.1109/TMM.2024.3359769, 2024.
[27] K. Xu et al., "Stage-by-stage Wavelet Optimization Refinement Diffusion Model for Sparse-View CT Reconstruction," IEEE Trans. Med. Imaging, DOI: 10.1109/TMI.2024.3355455, 2024.
[28] H. Phung et al., "Wavelet Diffusion Models Are Fast and Scalable Image Generators," Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10199–10208, Vancouver, BC, Canada, 2023.
[29] N. B. Bynagari, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Asian Journal of Applied Sciences & Eng., vol. 8, no. 1, pp. 25–34, Apr. 2019.
[30] S. Barratt and R. Sharma, "A Note on the Inception Score," Proc. ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Stockholm, Sweden, [Online], Available https://drive.google.com/file/d/1y--gMfGtjlNcQXTPm-vLhuqlUbh19nF9/view, 2018.

ملخص البحث:

تستند غالبية تقنيات تركيب بَصَمات الأصابع على استخدام الشّبكات التّوليدية التّنافسية (GANs). وحديثاً، أثبت النّموذج الاحتمالي للانتشار القائم على إزالةِ الضّجيج (DDPM) أنّه أكثرُ فاعليةً من ذلك النّوع من الشّبكات في العديد من السّيناريوهات، لا سيما فيما يتعلّق بالتنوُّع والوُضوح.

في هذا البحث، نعملُ على تطوير نموذجٍ محسَّنٍ يقومُ على تقنية (DDPM) لتوليد بَصَمات الأصابع. وبالتحديد، فإنّه يتمّ تفكيك الصّورة إلى صُوَر فرعية بنطاقاتٍ تردُّديةٍ فرعيةٍ مختلفة، وذلك باستخدام طريقة تحويل حُزَم المويْجات (WPT) التي تجعل تقنية (DDPM) تعمل عند مستوى أكثر محلّيةً وتفصيلاً، مُسترجعةً بذلك خصائص البيانات على نحوٍ دقيقٍ. من ناحية أخرى، تمّ استخدام تقنيةٍ جديدةٍ للتّعامل مع الضّجيج لتحلَّ محلّ الاستراتيجية الخطّية لمعالجة الضّجيج، بحيث تكون عملية التّعامُل مع الضّجيج أكثر سلاسة.

وقد برهنت التّجاربُ على عددٍ من مَجْموعات البيانات المتعلّقة ببَصَمات الأصابع أن النّموذج المقترح في هذا البحث يتفوَّقُ على عددٍ من النّماذج المشابهة الواردة في أدبيات الموضوع.