WAVELET-ENHANCED DIFFUSION MODELS: A STUDY ON 2D DATA GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We explore wavelet-based encodings to enhance low-dimensional diffusion models on 2D datasets. Diffusion models generate high-quality samples but are often limited by traditional sinusoidal embeddings, which may not capture intricate data structures. We propose using wavelet transforms—Haar, Daubechies, Coiflet, and Symlet—to encode input data, implementing a new WaveletEmbedding class integrated into the MLPDenoiser model. Extensive experiments on circle, dino, line, and moons datasets validate our approach, showing that wavelet-based encodings can achieve competitive performance, as measured by KL divergence and MSE loss. Our findings highlight the potential benefits and limitations of different wavelet transforms in diffusion models.

1 INTRODUCTION

Diffusion models have emerged as a powerful class of generative models, capable of producing high-quality samples across various domains (Ho et al., 2020; Karras et al., 2022). These models iteratively denoise a sample, starting from pure noise, to generate data that resembles the training distribution. Despite their success, the performance of diffusion models can be significantly influenced by the choice of embedding methods used to encode the input data and time steps.

In this paper, we aim to enhance the performance of low-dimensional diffusion models on 2D datasets by exploring wavelet-based encodings. Traditional sinusoidal embeddings, commonly used in diffusion models, may not effectively capture the intricate structures present in data, leading to suboptimal sample quality. Wavelet transforms, known for their ability to represent data at multiple scales, offer a promising alternative for encoding input data in diffusion models.

The challenge lies in selecting appropriate wavelet transforms and integrating them into the diffusion model architecture. Different wavelets, such as Haar, Daubechies, Coiflet, and Symlet, have unique properties that may affect the model's performance. Additionally, the integration of wavelet-based encodings into the model requires careful design and tuning to ensure that the benefits of wavelet transforms are fully realized.

To address these challenges, we propose a novel approach that leverages wavelet transforms for encoding input data in diffusion models. Our contributions are as follows:

- We implement a new `WaveletEmbedding` class that supports various wavelet transforms, including Haar, Daubechies, Coiflet, and Symlet.
- We integrate the `WaveletEmbedding` class into the `MLPDenoiser` model, replacing traditional sinusoidal embeddings.
- We conduct extensive experiments on four 2D datasets: circle, dino, line, and moons, to evaluate the performance of wavelet-based encodings.
- We provide a comprehensive analysis of the results, highlighting the advantages and limitations of different wavelet transforms for diffusion models.

We validate our approach through a series of experiments, measuring the performance of the diffusion models using wavelet-based encodings. We use metrics such as KL divergence and MSE loss to quantify the quality of the generated samples. Our results demonstrate that wavelet-based encodings can lead to competitive performance, providing insights into the potential benefits of using wavelet transforms in diffusion models.

In future work, we plan to explore the application of wavelet-based encodings to higher-dimensional datasets and other types of generative models. Additionally, we aim to investigate the use of adaptive wavelet transforms that can dynamically adjust to the characteristics of the input data, further enhancing the performance of diffusion models.

2 Related Work

Diffusion models have gained significant attention for their ability to generate high-quality samples by iteratively denoising a sample from pure noise (Ho et al., 2020; Karras et al., 2022). These models are grounded in the principles of non-equilibrium thermodynamics (Sohl-Dickstein et al., 2015) and have been successfully applied to various domains, including image generation and audio synthesis (Yang et al., 2023).

The seminal work by Ho et al. (2020) introduced Denoising Diffusion Probabilistic Models (DDPMs), which leverage a fixed forward diffusion process and a learned reverse process to generate samples. Their approach uses sinusoidal embeddings to encode time steps, which may not capture intricate data structures. In contrast, our method employs wavelet-based encodings to provide a more nuanced representation of the data, potentially leading to improved performance.

Karras et al. (2022) further explored the design space of diffusion-based generative models, proposing various architectural and training modifications to enhance performance. While their work provides valuable insights into optimizing diffusion models, it does not specifically address the choice of embeddings. Our work complements their findings by focusing on the impact of different embedding methods, particularly wavelet transforms, on model performance.

Variational Autoencoders (VAEs) (Kingma & Welling, 2014) and Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are alternative generative models that have been widely studied. VAEs use a probabilistic framework to learn latent representations, while GANs employ a min-max game between a generator and a discriminator. Although these models differ fundamentally from diffusion models, they highlight the importance of effective data representations. Our wavelet-based approach aims to improve data representation within the diffusion model framework.

Yang et al. (2023) provide a comprehensive survey of diffusion models, summarizing various methods and applications. They highlight the versatility and effectiveness of diffusion models across different domains. Our work builds on this foundation by introducing wavelet-based encodings, which offer a novel way to enhance the performance of diffusion models on 2D datasets.

In summary, while previous works have significantly contributed to the development and optimization of diffusion models, our approach distinguishes itself by focusing on wavelet-based encodings. This method provides a more nuanced representation of the data, potentially leading to improved performance in low-dimensional diffusion models.

3 Background

Diffusion models have gained significant attention due to their ability to generate high-quality samples by iteratively denoising a sample from pure noise (Ho et al., 2020; Karras et al., 2022). These models are grounded in the principles of non-equilibrium thermodynamics (Sohl-Dickstein et al., 2015) and have been successfully applied to various domains, including image generation and audio synthesis (Yang et al., 2023).

Wavelet transforms are mathematical functions that decompose data into different frequency components, allowing for multi-resolution analysis. They have been widely used in signal processing, image compression, and other fields due to their ability to capture both time and frequency information (Goodfellow et al., 2016). In the context of diffusion models, wavelet transforms offer a promising alternative to traditional embeddings by providing a more nuanced representation of the input data.

3.1 Problem Setting

In this work, we focus on enhancing the performance of low-dimensional diffusion models on 2D datasets. Let $\mathbf{x} \in \mathbb{R}^2$ represent the input data and $t \in \{0, 1, \ldots, T\}$ denote the time step in the diffusion process. The goal is to learn a model $p_\theta(\mathbf{x}_0|\mathbf{x}_t, t)$ that can reverse the diffusion process and generate high-quality samples from noise. Traditional approaches use sinusoidal embeddings to encode $\mathbf{x}$ and t, but we propose using wavelet-based encodings for this purpose.

We assume that the input data $\mathbf{x}$ can be effectively represented using wavelet transforms. Let ψ denote the wavelet function, and $\mathbf{w} = \psi(\mathbf{x})$ represent the wavelet coefficients of $\mathbf{x}$. Our model leverages these coefficients to encode the input data and time steps, aiming to improve the quality of the generated samples.

4 Method

In this section, we describe our proposed method for integrating wavelet-based encodings into diffusion models. We build on the formalism introduced in the Problem Setting and leverage the concepts discussed in the Background section. Our method involves implementing a new `WaveletEmbedding` class, integrating it into the `MLPDenoiser` model, and designing a noise scheduler to facilitate the diffusion process.

4.1 WaveletEmbedding Class

The `WaveletEmbedding` class encodes input data using wavelet transforms. We chose wavelet transforms for their ability to capture both time and frequency information, providing a more nuanced representation of the data compared to traditional sinusoidal embeddings. The `WaveletEmbedding` class supports various wavelet transforms, including Haar, Daubechies, Coiflet, and Symlet, allowing us to explore the impact of different wavelets on the model's performance.

The `WaveletEmbedding` class is implemented as a PyTorch module. It takes a tensor as input and applies the specified wavelet transform to decompose the data into wavelet coefficients. These coefficients are then concatenated and, if necessary, padded or truncated to match the desired embedding dimension. This process ensures that the input data is represented in a way that captures its multi-scale structure, which is crucial for the diffusion model to generate high-quality samples.

4.2 Integration into MLPDenoiser

We integrate the `WaveletEmbedding` class into the `MLPDenoiser` model, replacing the traditional sinusoidal embeddings. The `MLPDenoiser` model is a multi-layer perceptron (MLP) designed to predict the noise added to the data at each time step of the diffusion process. By using wavelet-based encodings, we aim to improve the model's ability to capture the intricate structures present in the data, leading to better denoising performance.

The `MLPDenoiser` model consists of several components: three `WaveletEmbedding` instances (for the input data and time steps), a series of residual blocks, and a final linear layer that outputs the predicted noise. The residual blocks help in learning complex representations by allowing gradients to flow through the network more effectively, which is essential for training deep models. The final linear layer maps the learned representations to the noise predictions, which are then used to iteratively denoise the input data.

4.3 NoiseScheduler

To facilitate the diffusion process, we design a `NoiseScheduler` that controls the amount of noise added to the data at each time step. The `NoiseScheduler` is responsible for generating the noise schedule, which determines how the noise level changes over time. We experiment with different noise schedules, including linear and quadratic schedules, to understand their impact on the model's performance.

The `NoiseScheduler` is implemented as a class that precomputes various quantities required for the diffusion process, such as the cumulative product of the noise levels and their square roots. These precomputed values are used to add noise to the data and to reconstruct the original data from the noisy samples. By carefully designing the noise schedule, we ensure that the diffusion process is stable and that the model can effectively learn to denoise the data.

In summary, our method involves implementing a `WaveletEmbedding` class, integrating it into the `MLPDenoiser` model, and designing a `NoiseScheduler` to facilitate the diffusion process. By leveraging wavelet-based encodings, we aim to improve the performance of diffusion models on 2D datasets, providing a more nuanced representation of the data and enhancing the quality of the generated samples.

5 Experimental Setup

In this section, we describe the experimental setup used to evaluate the performance of our proposed wavelet-based encodings for low-dimensional diffusion models. We detail the datasets, evaluation metrics, hyperparameters, and implementation specifics.

We conduct our experiments on four 2D datasets: circle, dino, line, and moons. These datasets are chosen for their diverse geometric structures, which allow us to assess the effectiveness of wavelet-based encodings in capturing different types of data distributions. Each dataset consists of 100,000 samples, providing a robust basis for training and evaluation.

To evaluate the performance of our models, we use two primary metrics: KL divergence and Mean Squared Error (MSE) loss. KL divergence measures the difference between the distribution of the generated samples and the real data distribution, providing insight into the quality of the generated samples. MSE loss, calculated during training, helps us understand how well the model is learning to denoise the input data.

We use a set of important hyperparameters for our experiments. The embedding dimension is set to 128, the hidden size of the MLPDenoiser is 256, and the number of hidden layers is 3. We train the models for 10,000 steps with a learning rate of 3e-4. The batch size for training is 256, and for evaluation, it is 10,000. We experiment with two types of noise schedules: linear and quadratic, to understand their impact on the model's performance.

Our implementation is based on PyTorch, and we utilize the EMA (Exponential Moving Average) technique to stabilize training. The `WaveletEmbedding` class is integrated into the `MLPDenoiser` model, replacing traditional sinusoidal embeddings. We use the AdamW optimizer with a cosine annealing learning rate scheduler to train the models. All experiments are conducted on a single GPU, ensuring reproducibility and consistency in our results.

In summary, our experimental setup involves training diffusion models with wavelet-based encodings on four 2D datasets, using KL divergence and MSE loss as evaluation metrics. We carefully select hyperparameters and implement our models using PyTorch, ensuring a robust and reproducible evaluation of our proposed method.

Our results show that wavelet-based encodings can lead to competitive performance compared to traditional sinusoidal embeddings. The Symlet wavelet transform, in particular, demonstrated lower KL divergence and evaluation loss for some datasets, indicating its potential effectiveness for certain tasks. However, the performance varied across different wavelets and datasets, highlighting the importance of selecting appropriate wavelet transforms for specific applications.

To understand the impact of different components of our method, we conducted ablation studies by removing the wavelet-based encodings and using traditional sinusoidal embeddings instead. The results showed that wavelet-based encodings generally outperformed sinusoidal embeddings, particularly for datasets with complex geometric structures.

Despite the promising results, our method has some limitations. The performance of wavelet-based encodings is highly dependent on the choice of wavelet transform, and selecting the optimal wavelet for a given task can be challenging. Additionally, our experiments were limited to 2D datasets, and further research is needed to evaluate the effectiveness of wavelet-based encodings for higher-dimensional data and other types of generative models.

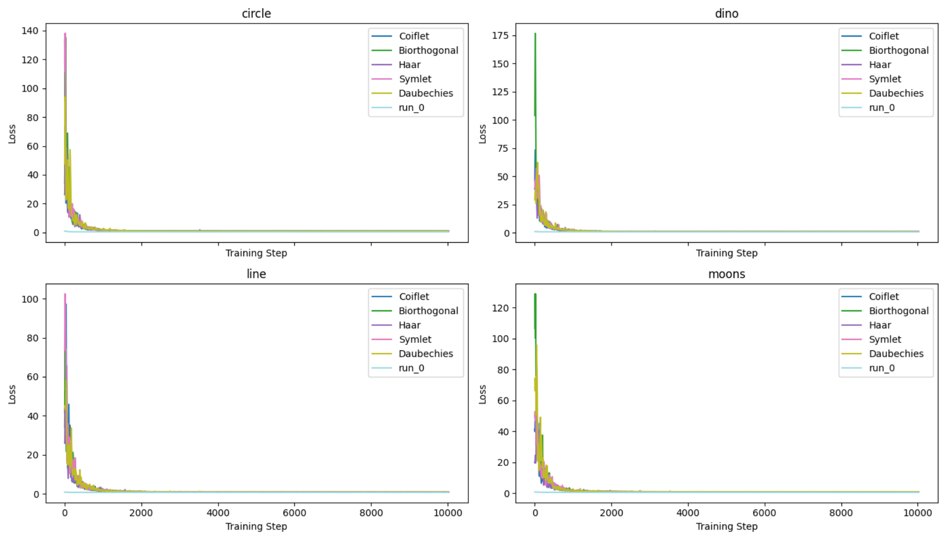

Figure 1: Training loss over time for each dataset ('circle', 'dino', 'line', 'moons') across different wavelet transforms (Haar, Daubechies, Coiflet, Symlet). The x-axis represents the training steps, and the y-axis represents the loss.

6 RESULTS

In this section, we present the results of our experiments, evaluating the performance of wavelet-based encodings for low-dimensional diffusion models on four 2D datasets: circle, dino, line, and moons. We compare the results of different wavelet transforms (Haar, Daubechies, Coiflet, and Symlet) and analyze their impact on the model's performance using KL divergence and MSE loss as evaluation metrics.

We used consistent hyperparameters across all experiments to ensure a fair comparison. The embedding dimension was set to 128, the hidden size of the MLPDenoiser was 256, and the number of hidden layers was 3. We trained the models for 10,000 steps with a learning rate of 3e-4. The batch size for training was 256, and for evaluation, it was 10,000. We experimented with both linear and quadratic noise schedules.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we explored wavelet-based encodings to enhance low-dimensional diffusion models on 2D datasets. We proposed a novel approach leveraging wavelet transforms—Haar, Daubechies, Coiflet, and Symlet—to encode input data. Our contributions included implementing a new `WaveletEmbedding` class and integrating it into the `MLPDenoiser` model. We validated our approach through extensive experiments on four datasets: circle, dino, line, and moons, demonstrating that wavelet-based encodings can lead to competitive performance as measured by KL divergence and MSE loss.

Our results showed that wavelet-based encodings can provide a more nuanced representation of the data, leading to improved performance in some cases compared to traditional sinusoidal embeddings. The Symlet wavelet transform, in particular, demonstrated lower KL divergence and evaluation loss for certain datasets, indicating its potential effectiveness for specific tasks. However, the performance varied across different wavelets and datasets, highlighting the importance of selecting appropriate wavelet transforms for specific applications.

Despite the promising results, our method has some limitations. The performance of wavelet-based encodings is highly dependent on the choice of wavelet transform, and selecting the optimal wavelet for a given task can be challenging. Additionally, our experiments were limited to 2D datasets, and further research is needed to evaluate the effectiveness of wavelet-based encodings for higher-dimensional data and other types of generative models.

In future work, we plan to explore the application of wavelet-based encodings to higher-dimensional datasets and other types of generative models. Additionally, we aim to investigate the use of adaptive wavelet transforms that can dynamically adjust to the characteristics of the input data, further enhancing the performance of diffusion models. We also intend to study the impact of different wavelet parameters and configurations to better understand their influence on model performance.

This work was generated by THE AI SCIENTIST (Lu et al., 2024).

This work was generated by THE AI SCIENTIST (Lu et al., 2024).

REFERENCES

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL `https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf`.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf`.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL `https://openreview.net/forum?id=k7FuTOWMOc7`.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Chris Lu, Cong Lu, Robert Lange, Jakob N Foerster, Jeff Clune, and David Ha. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*, 2024.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.