Smoothed Energy Guidance: Guiding Diffusion Models with Reduced Energy Curvature of Attention.

ML Reproducibility Challenge-2025

Anonymous authors
Paper under double-blind review

Abstract

This study is part of the MLRC Reproducibility Challenge 2025, aiming to reproduce and improve the results from a NeurIPS 2024 submission *Smoothed Energy Guidance (SEG): Guiding Diffusion Models with Reduced Energy Curvature of Attention.* The work proposed in the SEG paper faced key limitations, including the lack of an ablation study for optimal kernel size selection and unexplored alternative blurring strategies within diffusion models, which could offer valuable insights into enhancing image quality and model robustness. Furthermore, the approach employed unnecessary smoothing throughout all iterations of the denoising process, which not only diminished the clarity of the output but also resulted in increased computational costs. To address these issues, we conducted a detailed ablation study and explored more efficient alternatives, including Exponential Moving Average (EMA) and BoxBlur using integral images, to improve computational efficiency while maintaining image quality. Our findings provide insights into optimizing smooth energy guidance in diffusion models, reducing computational overhead while improving image quality. Our code is available at SEG Reproducibility Challenge 2025 Repository.

1 Introduction

Smoothed Energy Guidance (SEG) was proposed as a tuning- and condition-free method for improving self-attention mechanisms from an energy-based perspective. Instead of relying on a guidance scale parameter, SEG directly blurs attention weights using a Gaussian kernel Gonzalez-Jimenez et al. (2021) to reduce the curvature of the underlying energy function. This approach enables continuous control over energy modulation while avoiding the side effects of large guidance scales. A novel query blurring technique was proposed in the SEG paper to achieve full attention weight smoothing without quadratic cost using Eq.1.

$$\mathrm{SEG}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathcal{B}(\mathbf{Q})\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \tag{1}$$

where $\mathcal{B}$ is the Toeplitz matrix for Gaussian blur.

Furthermore, the SEG paper revised the Stochastic Differential Equation (SDE) for Diffusion Models by incorporating energy guidance, which leads to improved image generation quality Nichol & Dhariwal (2021). Eq. 2 is the updated formulation:

$$d\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g(t)^2\left((1 - \gamma_{\text{cfg}} + \gamma_{\text{seg}})s_\theta(\mathbf{x}, t) + \gamma_{\text{cfg}}s_\theta(\mathbf{x}, t, c) - \gamma_{\text{seg}}\tilde{s}_\theta(\mathbf{x}, t)\right)\right]dt + g(t)d\bar{\mathbf{w}} \tag{2}$$

where s_θ is the score model and $\tilde{s}_\theta$ is the score model calculated using blurred attention weights. The parameters γ_{cfg} and γ_{seg} control the guidance strength of classifier-free guidance Ho & Salimans (2022) and energy-based guidance, respectively Wang (2023).

1.1 Experimental Setup and Code

What was easy: The codebase provided by the authors of the SEG paper is well-documented, making it easy to conduct experiments with varying blur strength (σ). To enhance clarity and maintainability, we have modularized the codebase into distinct components, separating key conceptual elements such as the model pipeline, blurring strategies, and attention_processor modules.

What was difficult: The original code was constrained to studying the effects of Gaussian blurring with varying σ and guidance strength (γ_{seg}), limiting flexibility in experimentation. To enhance the reproducibility and efficiency of the code, we optimized the 2D Gaussian blur convolution by caching the Gaussian kernel, significantly reducing computational overhead. Additionally, we introduced flexible hyperparameter settings, allowing adjustments to σ, kernel size, blurring schedules, and selective blurring of attention layers. Furthermore, we expanded the model pipeline to support alternative, computationally efficient smoothing strategies beyond Gaussian convolution, enabling broader exploration of image generation quality improvements. To quantitatively assess the impact of blurring, we incorporated various metrics like Frobenius Norm Böttcher & Wenzel (2008), Laplacian Variance Bansal et al. (2016), and Gradient Entropy Zhao et al. (2016) applied to attention layers of the mid-block of the U-Net of the diffusion model Rombach et al. (2022).

1.2 Scope of Reproducibility

This investigation focuses on several critical aspects concerning the reproducibility of the SEG paper:

- **Analysis of Kernel Size Variations in Energy-Guided Generation**: The influence of energy guidance on the quality of image generation, particularly regarding the application of varying Gaussian blur kernel sizes, has not been thoroughly analyzed or discussed.
- **Impact of Gaussian Blur on Different Attention Layers in U-Net**: There is ambiguity regarding the specific set of attention layers (down, mid, up) within the U-Net architecture, to which part of the U-Net Gaussian blur should be applied, aiming to smooth the energy landscape and enhance the quality of image generation.
- **Minimizing Redundant Computations in Blurring Iteratively**: The application of blurring at each iteration to smooth the energy landscape results in unnecessary computations, which could potentially be eliminated to improve efficiency.

2 Experiments

To validate the claims outlined in the scope of reproducibility, we conducted rigorous experiments using a seed value of 77 to systematically analyze the impact of key factors on image generation quality. The detailed study using different seed values is provided in the appendix section for reference.

2.1 Analysis of Kernel Size Variations in Energy-Guided Generation

The SEG paper defines the kernel size as a function of σ using the following relation:

$$\text{kernel size} = \lceil 6\sigma \rceil + 1 - (\lceil 6\sigma \rceil \bmod 2) \tag{3}$$

For σ=10, the given formula produces a kernel size 61, which is computationally impractical. A series of experiments were conducted to evaluate the influence of kernel size and to identify the optimal kernel value on the quality of image generation within the framework of energy-guided diffusion. Figure 1 presents images generated using different kernel sizes. The results indicate that kernel sizes of 31 yield almost similar image quality as obtained in the SEG paper. Our investigation further reveals that increasing the kernel size beyond 31 leads to minor improvements in image quality while significantly escalating computational costs.

2.2 Impact of Gaussian Blur on Different Attention Layers in U-Net

The U-Net part of the latent diffusion models typically includes multiple attention layers in their down, mid, and up regions. We systematically evaluated the impact of applying Gaussian blur to each region, which has not been reported in the SEG paper. The effect of blurring at different attention layers of U-Net is shown in Figure 2.:

Figure 1: Image quality assessment using Gaussian blur with varying kernel sizes, where σ=10. Blurring was applied to the mid-attention layers and during the initial 30-35% of total iterations.

- **Down and UP Attention Layers**: Blurring at the down and up regions of the U-Net resulted in highly distorted and unnatural images. The color and texture appeared scrambled, with a significant loss of structure and coherence. Instead of smooth transitions or meaningful abstractions, the images exhibited disorganized patterns and blocky artifacts.

- **Mid Attention Layers**: The middle images in the second row, which appear more natural than the others, suggest that blurring in the mid-layer impacts the overall structure without completely disrupting image coherence. While textures are slightly smoothed, the images still preserve recognizable forms and objects.

The observations from the images confirm that selective blurring affects different regions of U-Net differently. Applying it at the mid-region is the most balanced approach, while blurring in the down or up regions severely degrades image quality.

2.3 Minimizing Redundant Computations in Blurring Iteratively

Applying blurring at each iteration to smooth the energy landscape leads to unnecessary computations, which can be optimized for improved efficiency. Figure 2 illustrates the image quality achieved when blurring is applied at different stages of the iteration process.

- In the SEG paper, Gaussian blurring was applied at every iteration, significantly increasing computational costs. However, our experiments reveal that applying Gaussian blur only during the first 30–35% of the reverse diffusion steps is sufficient to guide the model toward stable solutions.

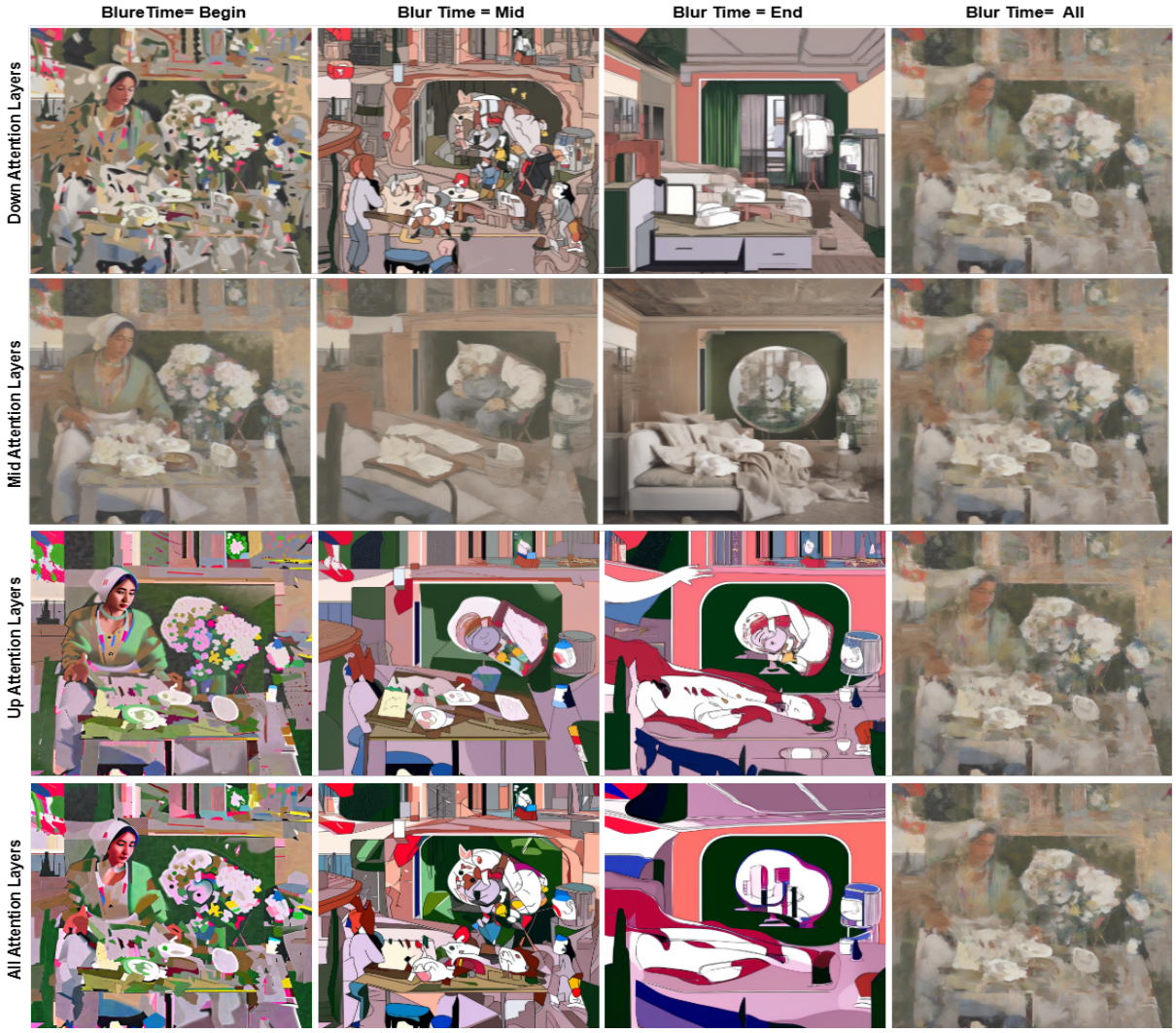


Figure 2: Image quality generation for a fixed kernel size of 31 and $\sigma = 10$. The 30 iterations of reverse diffusion are divided into three equal phases: Begin (iterations 1-10), Mid (iterations 11-20), and End (iterations 21-30).

- This finding suggests that smoothed energy guidance is most effective in the initial stages of the reverse diffusion process, ensuring a stable generation trajectory without altering the final output.

2.4 Experiment Beyond Paper

In this section, we interpret the SDE equation of diffusion models through the lens of deep learning optimizers. We provide a mathematically equivalent perspective on the linear combination of different score models $\mathbf{s}_\theta$ as described in Eq. 2.

$$d\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \mathbf{s}^\star_\theta(\mathbf{x}, t)\right] dt + g(t) d\bar{\mathbf{w}} \tag{4}$$

where

$$\mathbf{s}^\star_\theta(\mathbf{x}, t) = (1 - \gamma_{\text{cfg}} + \gamma_{\text{seg}})\mathbf{s}_\theta(\mathbf{x}, t) + \gamma_{\text{cfg}}\mathbf{s}_\theta(\mathbf{x}, t, c) - \gamma_{\text{seg}}\tilde{\mathbf{s}}_\theta(\mathbf{x}, t) \tag{5}$$

$$= \mathbf{s}_\theta(\mathbf{x}, t) + \gamma_{\text{cfg}}(\mathbf{s}_\theta(\mathbf{x}, t, c) - \mathbf{s}_\theta(\mathbf{x}, t)) + \gamma_{\text{seg}}(\mathbf{s}_\theta(\mathbf{x}, t) - \tilde{\mathbf{s}}_\theta(\mathbf{x}, t)) \tag{6}$$

In Eq. 6, $\mathbf{s}_\theta(\mathbf{x}, t)$ represents the unguided score model. The term $(\mathbf{s}_\theta(\mathbf{x}, t, c) - \mathbf{s}_\theta(\mathbf{x}, t))$ accounts for classifier guidance, while $(\tilde{\mathbf{s}}_\theta(\mathbf{x}, t) - \mathbf{s}_\theta(\mathbf{x}, t, c))$ represents energy guidance. The final resultant score model is denoted as $\mathbf{s}^\star_\theta(\mathbf{x}, t)$.

Assuming $\gamma_{\text{cfg}} = 0$, we obtain:

$$\mathbf{s}^\star_\theta(\mathbf{x}, t) = \mathbf{s}_\theta(\mathbf{x}, t) + \gamma_{\text{seg}}(\tilde{\mathbf{s}}_\theta(\mathbf{x}, t) - \mathbf{s}_\theta(\mathbf{x}, t, c)) \tag{7}$$

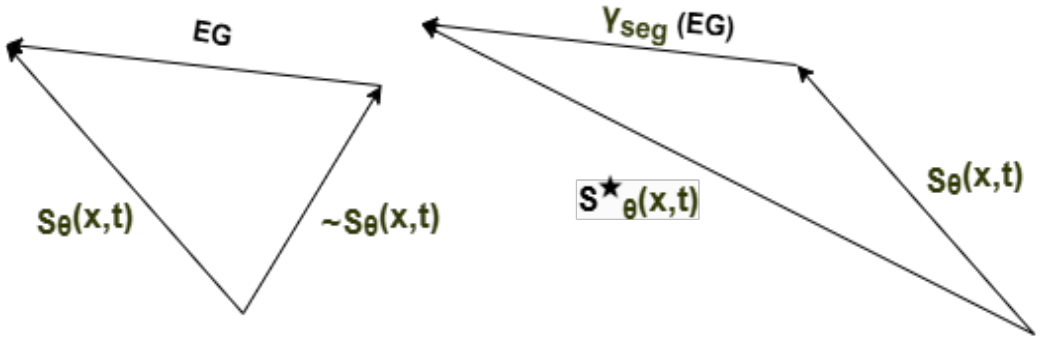


Figure 3: Illustration of energy guidance in diffusion models, drawing an analogy to momentum-based optimizers like Adam, where γ_{seg} influences the update direction similar to accumulated momentum, leading to a smoother optimization landscape

Figure 3 suggests that energy guidance functions similarly to momentum in optimizers like Adam, where the final gradient incorporates both the current gradient and accumulated momentum, leading to a smoother loss landscape. For extreme blur strength ($\sigma \to \infty$), generation quality improves, as demonstrated in the SEG paper. In this scenario, maximal information destruction occurs, shifting from weighted averaging via a Gaussian kernel to simple averaging, which increases the orthogonality between $\mathbf{s}_\theta$ and $\tilde{\mathbf{s}}_\theta$. This enhanced orthogonality plays a key role in improving generation quality.

Given that the maximum benefit of orthogonality is observed at $\sigma \to \infty$, we explore alternative smoothing techniques that achieve similar effects but are computationally far more efficient than Gaussian blurring with large kernel sizes. Below, we introduce two blurring techniques—Exponential Moving Average (EMA) and Box Blur- which offer a more computationally efficient approach to smoothing energy landscapes in diffusion models. These mathematical insights lead us to search for various smoothing techniques widely used in deep learning, like EMA and BoxBlur, via integral images.

2.4.1 Exponential Moving Average Smoothing

EMA smoothing is a technique used to smooth data by giving more weight to recent observations while still considering past values. It is commonly used in time series analysis, signal processing, and deep learning for stabilizing updates.

The EMA at time t is computed as:

$$\text{EMA}_t = (1 - \beta)x_t + \beta\text{EMA}_{t-1} \tag{8}$$

where x_t is the current data point, β is the smoothing factor ($0 < \beta \leq 1$) increasing linearly, and EMA_{t-1} represents the previous EMA value.

2.4.2 Box Blur (BoxBlur)

Box Blur (also known as a mean filter) is a smoothing technique where each pixel in an image is replaced by the average of the pixels in its surrounding neighborhood. It is commonly used in image processing to reduce noise and create a blurring effect. For a given image I and a kernel of size $k \times k$, the Box Blur is computed as:

$$I'_{(x,y)} = \frac{1}{k^2} \sum_{i=-\frac{k}{2}}^{\frac{k}{2}} \sum_{j=-\frac{k}{2}}^{\frac{k}{2}} I(x+i, y+j) \tag{9}$$

where $I'_{(x,y)}$ is the blurred pixel value at location (x, y), k is the kernel size (typically an odd number), and $I(x+i, y+j)$ are the neighboring pixel values within the kernel. We linearly inerpolate the original image with the blurred image using Eq. 10

$$Img'(x, y) = (1 - \alpha) \cdot I(x, y) + \alpha \cdot I'(x, y) \tag{10}$$

where Img' is the final smoothed image.

2.5 Quantitative Assessment Through Metrics

The following metrics were utilized to quantitavely assess the impact of blurring on image generation.

Frobenius Norm: The Frobenius norm provides a scalar value representing the "energy" of a matrix, analogous to the Euclidean norm for vectors. For a matrix A of size $m \times n$, the Frobenius norm is defined as:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} \tag{11}$$

where a_{ij} represents the element at the ith row and jth column of matrix A.

A higher Frobenius norm indicates that the matrix has larger-valued elements, often implying greater variance, higher energy, or stronger signal intensity. Conversely, a lower Frobenius norm suggests smaller-valued elements, indicating lower energy, reduced variability, or smoother transitions within the matrix structure.

Laplacian Variance: A measure of image sharpness, where higher values indicate sharper images and lower values suggest smoothed images. It is computed using the Laplacian operator, which highlights edges through convolution with the 3×3 kernel:

$$K_{\text{Laplacian}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

For an image I, the Laplacian $L(I)$ is obtained by convolving I with $K_{\text{Laplacian}}$. The variance of $L(I)$ is given by:

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(L(I)_i - \mu)^2 \tag{12}$$

where $L(I)_i$ are Laplacian pixel values, μ is their mean, and N is the total pixels. Higher variance indicates sharper details.

Gradient Entropy: Gradient entropy measures the randomness in image gradients, reflecting texture complexity. Higher entropy indicates more detailed structures, while lower entropy suggests smoother regions.

- **Gradient Calculation**: Compute horizontal and vertical gradients using Sobel operators:

$$G_x = I * K_x, \quad G_y = I * K_y \tag{13}$$

  where I represents the input image, K_x and K_y are the Sobel kernels for computing horizontal and vertical gradients, respectively, and $*$ denotes the convolution operation.

- **Gradient Magnitude**: Calculate the overall gradient magnitude:

$$G = \sqrt{G_x^2 + G_y^2} \tag{14}$$

  where G_x and G_y represent the horizontal and vertical gradients.

- **Entropy Formula**: Measure the randomness of the gradient magnitudes:

$$H(G) = -\sum_i p_i \log_2(p_i) \tag{15}$$

  where $H(G)$ denotes the gradient entropy, p_i represents the probability of the i^{th} gradient magnitude value, and $\log_2$ refers to the logarithm to base 2.

Figure 4: Step-by-step image quality enhancement with linearly increasing β values, where β varies within a specified range. Blurring is applied using EMA smoothing.

3 Results and Discussion

This section explored the impact of different blurring techniques, including EMA and Box blur, on the image quality.

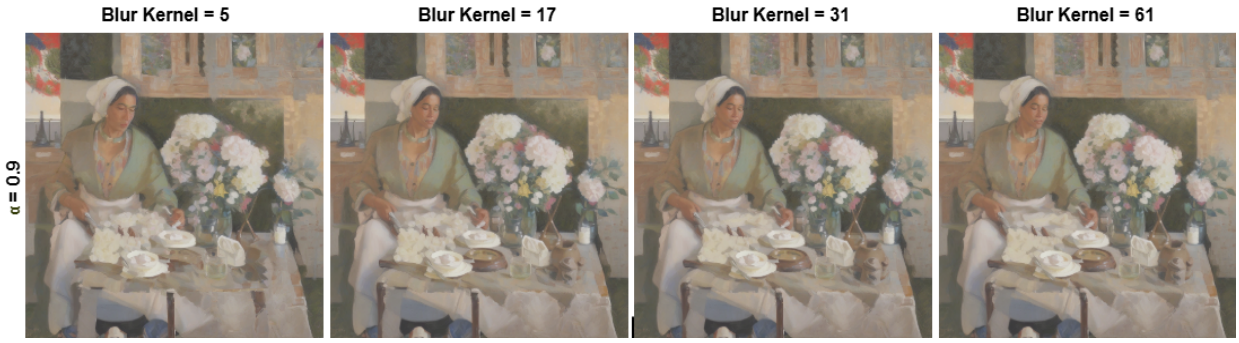


Figure 5: Image quality assessment using Box Blur with different kernel sizes.

Figure 6: Image quality evaluation using Box Blur with a kernel size of 31 and varying α values.

3.1 EMA and Box Blur Results

Figure 4 demonstrates that the optimal range for β in the EMA technique is close to 1, emphasizing the need for strong smoothing. In the early stages of reverse diffusion, skip connections and normalization layers tend to suppress blurring effects unless it is sufficiently strong. When applying high-intensity blurring with β nearly equal to 1, the final generated image exhibits significant improvements compared to those produced with lower β values or without energy guidance. The butterfly effect, discussed in section 3.2.1, highlights how heavily blurred attention maps initially exhibit only subtle differences that are not captured by the metrics used in this study. However, due to the expansive nature of the reverse diffusion process, the impact of strong blurring becomes increasingly significant in later diffusion stages.

The Box Blur technique has two key parameters: α and kernel size. Their optimal values were determined through multiple preliminary experiments. Figure 5 presents the variation in image quality with a fixed α value while varying the kernel size. It is observed that at lower kernel sizes, image quality deteriorates, whereas increasing the kernel size enhances the image clarity. Similarly, Figure 6 illustrates the effect of image quality when using a fixed kernel size and varying α. The results indicate that the image quality progressively improves as α increases.

Among the three blurring techniques, Gaussian Blur (from the SEG paper), EMA, and Box Blur effectively smoothen the energy landscape when **strong blurring** is applied. Table 1 illustrates the time complexity of the various blurring techniques. While comparing their computational complexity, the Box Blur and EMA emerged as more efficient alternatives to Gaussian Blur for achieving strong blurring.

Blurring Technique	Time Complexity	Optimal Parameter Ranges
Gaussian Blur	$O(T \times k^2)$	k = 31, $\sigma = 10$
EMA Smoothing	$O(T)$	$\beta = [0.95, 0.99]$
Box Blur	$O(T)$	$\alpha = 0.99$

Table 1: Comparison of Blurring Techniques: Time Complexity and Optimal Parameter Ranges. Here, T represents the token count in a query with the shape (batch, attention head, token, embedding dimension), while K denotes the kernel size

3.2 Impact of Query Blurring on Attention Mechanism: Analysis Using Gradient Entropy, Laplacian Variance, and Frobenius Norm

We analyze hidden states of shape (C,H,W) generated in the mid-regions of the U-Net's attention mechanism within the diffusion model. Our goal is to examine the effect of blurring the query and its influence on the final hidden states produced by the attention mechanism and subsequent normalization layers. Each metric is computed separately for each channel and then averaged. Figure 7 visualizes how these metrics vary across different blur settings throughout the reverse diffusion steps. As expected, gradient entropy and Laplacian variance decrease, indicating progressive denoising and stable image generation. In contrast, the Frobenius norm exhibits a different trend, initially dropping midway through the reverse diffusion process before rising again.

3.2.1 Amplification of Initial Perturbations (The Butterfly Effect)

In our experiments, we observed that the three global metrics: Frobenius Norm, Laplacian Variance, and Gradient Entropy measured on the hidden states were indistinguishable across different blurring settings for the first few iterations of the reverse diffusion process as seen in Figure 7. These metrics overlapped to the extent that differences were not detectable up to 13 decimal places. This substantial overlap indicates that the employed global metrics are not sensitive enough to capture the subtle local differences present when the noise level is very high in the initial iterations of reverse diffusion.

However, as the reverse diffusion process continues, the curves corresponding to different blurring settings begin to diverge sharply. This divergence is a direct consequence of reverse diffusion dynamics' nonlinear and iterative nature. Although no further blurring is applied after the initial stages, the minuscule differences introduced by distinct blurring configurations are preserved in the latent state and are gradually amplified over successive denoising iterations.

Mathematically, we model the reverse diffusion update at timestep t by the function

$$x_{t-1} = f_t(x_t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \, \epsilon_\theta(x_t, t) \right), \tag{16}$$

where α_t is determined by the noise schedule and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. Let $x_t^{(1)}$ and $x_t^{(2)}$ be two hidden states corresponding to different blurring conditions, differing by an infinitesimal perturbation

$$\delta_t = x_t^{(1)} - x_t^{(2)}, \tag{17}$$

with $\|\delta_t\|$ on the order of 10^{-13}. For sufficiently small δ_t, a first-order Taylor expansion of f_t around x_t yields

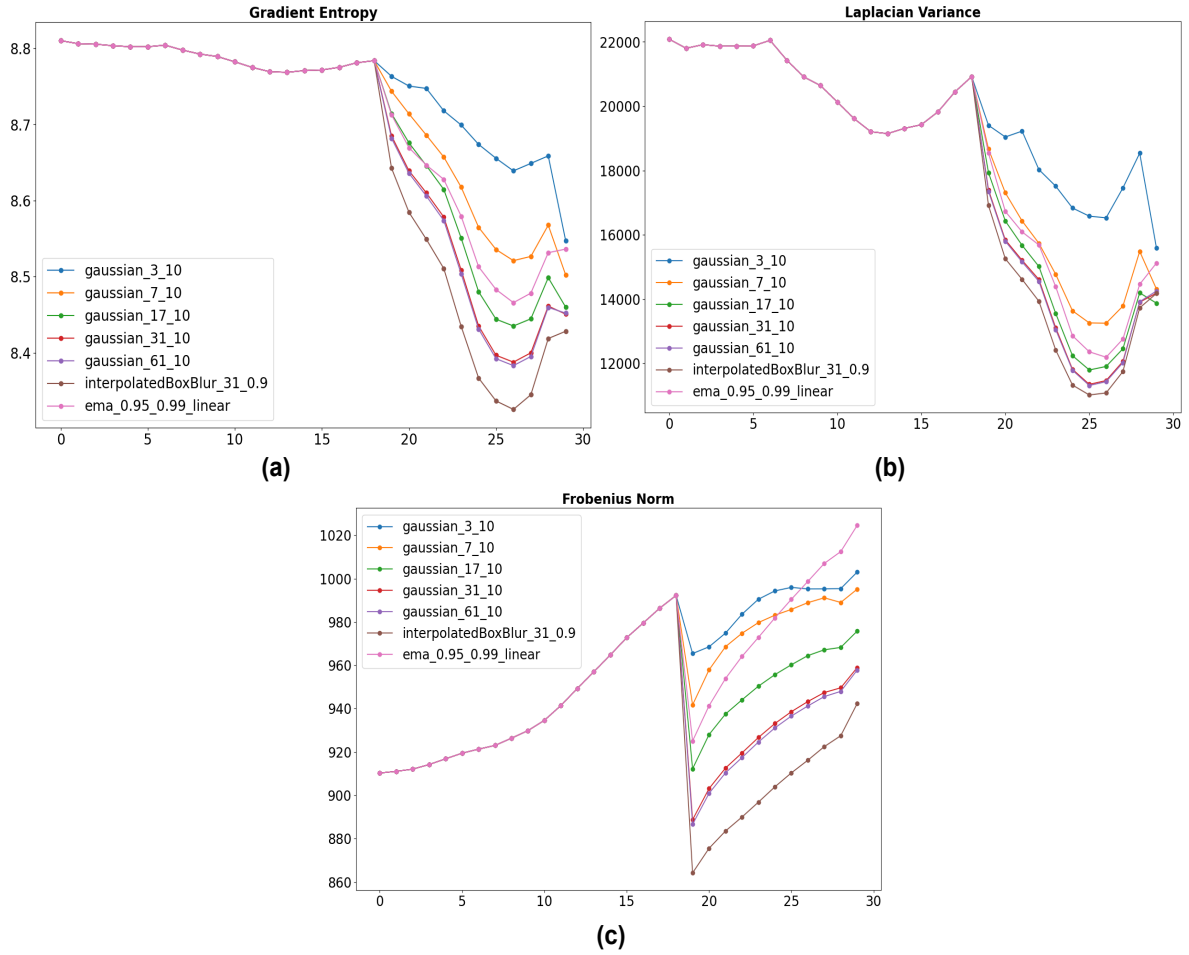
$$f_t(x_t + \delta_t) \approx f_t(x_t) + J_t \, \delta_t, \tag{18}$$

where the Jacobian $J_t = \frac{\partial f_t}{\partial x_t}$ quantifies the sensitivity of f_t to changes in x_t. Thus, the difference at timestep $t-1$ is approximated by

$$\delta_{t-1} \approx J_t \, \delta_t. \tag{19}$$

Iterating this process over k steps, the difference evolves as

$$\|\delta_{t-k}\| \approx \left\| \prod_{i=t-k+1}^{t} J_i \right\| \|\delta_t\|. \tag{20}$$



s

Figure 7: Variation of metrics like Gradient Entropy, Laplacian Variance and Frobenius Norm calculated on hidden states generated by attention layer of mid-block post attention mechanism and normalisation.

If the norm of the product of the Jacobians grows exponentially, i.e.,

$$\left\| \prod_{i=t-k+1}^{t} J_i \right\| \approx e^{\lambda k} \quad (\lambda > 0), \tag{21}$$

then even an initial perturbation of order 10^{-13} will be amplified as

$$\|\delta_{t-k}\| \approx e^{\lambda k} \|\delta_t\|. \tag{22}$$

This central equation captures the essence of the butterfly effect in the reverse diffusion process. Despite the fact that our global metrics initially fail to reveal any differences between the hidden states (showing a near-perfect overlap), the nonlinear dynamics — analogous to those in chaotic systems with a positive Lyapunov exponent — gradually amplify these minuscule differences. Consequently, the metric curves corresponding to different blurring settings diverge at later iterations, clearly demonstrating the expansive nature of the process.

A detailed analysis and derivation of this amplification mechanism is provided in Appendix A.

4 Conclusion

This study validates the findings of the SEG paper, demonstrating that energy smoothing through attention weight blurring enhances image generation quality while decoupling it from external guidance like CFG. Our analysis shows that subtle energy guidance minimally alters the generation trajectory, ensuring realistic outputs without artifacts. We further explored more efficient smoothing techniques, highlighting the importance of precise early-stage guidance. This approach not only improves generation quality but also suggests future research into its role in accelerating convergence during reverse diffusion, potentially reducing the required denoising steps.

References

Raghav Bansal, Gaurav Raj, and Tanupriya Choudhury. Blur image detection using laplacian operator and open-cv. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, pp. 63–67. IEEE, 2016.

Albrecht Böttcher and David Wenzel. The frobenius norm and the commutator. *Linear algebra and its applications*, 429(8-9):1864–1885, 2008.

Alvaro Gonzalez-Jimenez, Luca Lomazzi, Francesco Cadini, Alessio Beligni, Claudio Sbarufatti, Marco Giglio, and Andrea Manes. On the mitigation of the rapid algorithm uneven sensing network issue employing averaging and gaussian blur filtering techniques. *Composite Structures*, 278:114716, 2021.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Zihao Wang. Score-based generative modeling through backward stochastic differential equations: Inversion and generation. *arXiv preprint arXiv:2304.13224*, 2023.

Wenda Zhao, Zhijun Xu, and Jian Zhao. Gradient entropy metric and p-laplace diffusion constraint-based algorithm for noisy multispectral image fusion. *Information Fusion*, 27:138–149, 2016.

A Appendix

You may include other additional sections here.

A Detailed Mathematical Analysis of the Amplification Mechanism

In this appendix, we provide a detailed derivation of the exponential amplification of minute perturbations in the reverse diffusion process. This phenomenon—where global metrics initially fail to capture any differences (showing identical values up to 13 decimal places) and then diverge—illustrates a butterfly effect analogous to that observed in chaotic systems.

A.1 Reverse Diffusion Update Function

We define the reverse diffusion update at timestep t as

$$x_{t-1} = f_t(x_t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right), \tag{23}$$

where α_t is determined by the noise schedule and

$$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i,$$

and $\epsilon_\theta(x_t, t)$ denotes the noise prediction by the neural network.

A.2 Linearization via Taylor Expansion

Let $x_t^{(1)}$ and $x_t^{(2)}$ be two hidden states at time t corresponding to different blurring settings. Their difference is given by

$$\delta_t = x_t^{(1)} - x_t^{(2)}, \tag{24}$$

with $\|\delta_t\|$ on the order of 10^{-13}. For a sufficiently small δ_t, we can approximate the effect of f_t by a first-order Taylor expansion around $x_t^{(2)}$:

$$f_t(x_t^{(2)} + \delta_t) \approx f_t(x_t^{(2)}) + J_t\, \delta_t, \tag{25}$$

where the Jacobian J_t is defined as

$$J_t = \frac{\partial f_t}{\partial x_t}\Big|_{x_t = x_t^{(2)}}. \tag{26}$$

Thus, the difference at timestep $t-1$ becomes

$$\delta_{t-1} = f_t(x_t^{(1)}) - f_t(x_t^{(2)}) \approx J_t\, \delta_t. \tag{27}$$

A.3 Iterative Amplification of Perturbations

By applying the linearized update repeatedly over k iterations, the difference evolves as

$$\delta_{t-k} \approx \left(\prod_{i=t-k+1}^{t} J_i \right) \delta_t. \tag{28}$$

Taking the norm on both sides, we obtain

$$\|\delta_{t-k}\| \approx \left\| \prod_{i=t-k+1}^{t} J_i \right\| \|\delta_t\|. \tag{29}$$

In systems exhibiting sensitive dependence on initial conditions, the product of the Jacobian norms grows exponentially. That is, we can approximate

$$\left\| \prod_{i=t-k+1}^{t} J_i \right\| \approx e^{\lambda k}, \tag{30}$$

with $\lambda > 0$ acting as an effective Lyapunov exponent. Thus, we arrive at the central amplification equation:

$$\|\delta_{t-k}\| \approx e^{\lambda k} \|\delta_t\|. \tag{31}$$

A.4 Interpretation and Relation to Global Metrics

The derivation shows that, although the global metrics (L2 norm, Laplacian variance, and entropy) initially fail to capture any differences in the hidden states—reporting nearly identical values (up to 13 decimal places) across all blurring settings—the nonlinear dynamics of the reverse diffusion process amplify even infinitesimal differences. This exponential growth, captured by the effective Lyapunov exponent λ, explains the eventual divergence of the metric curves observed in later iterations. In essence, the system exhibits a butterfly effect, where the small, undetectable perturbations present in the early iterations become significantly amplified over time, leading to distinct final outcomes.

B Generalization Across Different Random Seeds

To further validate the robustness of our findings, we present additional qualitative results generated using the same experimental settings but with a different random seed (53647657).

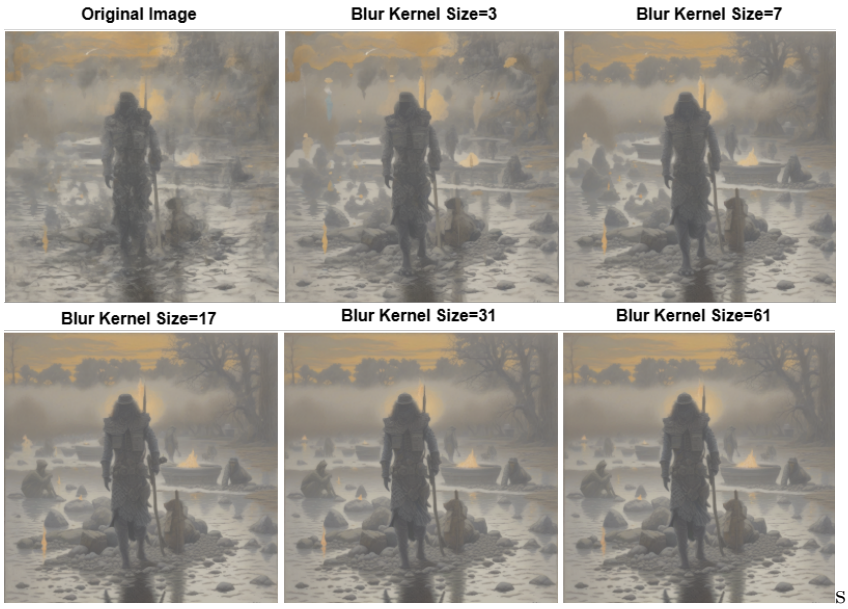


Figure 8: Image quality assessment using Gaussian blur with varying kernel sizes, where σ=10. Blurring was applied to the mid-attention layers and during the initial 30-35% of the total iterations with reproducibility Seed(53647657)

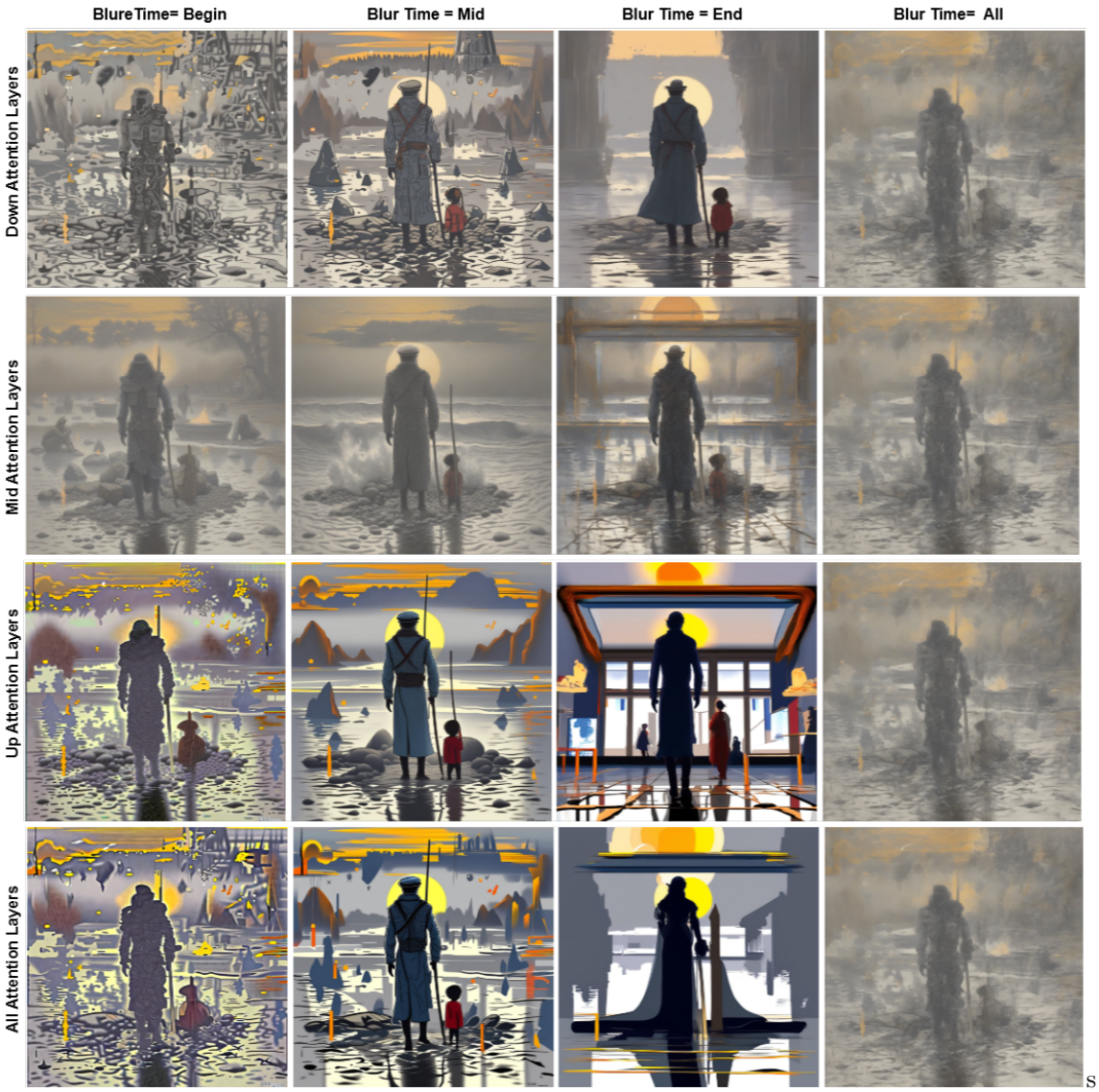


Figure 9: Image quality generation for a fixed kernel size of 31 and $\sigma = 10$. The 30 iterations of reverse diffusion are divided into three equal phases: Begin (iterations 1-10), Mid (iterations 11-20), and End (iterations 21-30), ensuring reproducibility with Seed (53647657).

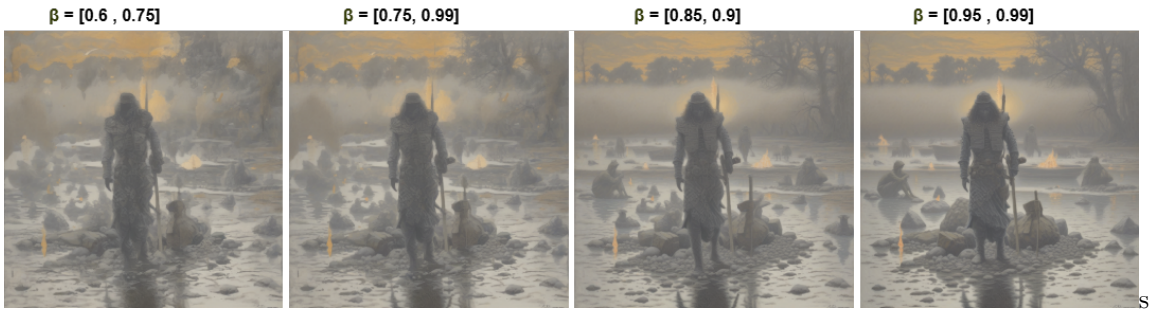


Figure 10: Figure 10: Step-by-step image quality enhancement with linearly increasing β values, where β varies within a specified range. Blurring is applied using EMA smoothing, with reproducibility Seed(53647657)

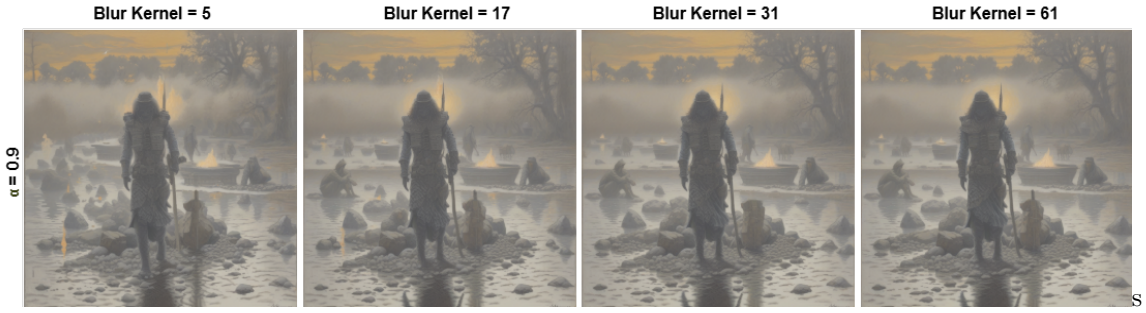


Figure 11: Figure 11: Image quality evaluation using Box Blur with a kernel size of 31 and varying α values, with reproducibility Seed(53647657)

Figure 12: Figure 12: Image quality evaluation using Box Blur with a kernel size of 31 and varying α values, with reproducibility Seed(53647657)