

Research Article

Enhancing GAN Training Stability for Image Super-Resolution Reconstruction with AdaBelief Optimization Strategy

Kun Cai , Jiancheng Cai , Hailong Li , Xiaofeng Wang , Chaochun Zhong* 

Electroacoustic Measurement Center, Guangzhou Institute of Measurement and Testing Technology, Guangzhou, China

Abstract

Generative Adversarial Networks (GANs) have achieved remarkable success in image super-resolution reconstruction, producing high-quality images from low-resolution inputs. However, their training process is often plagued by instability issues, such as mode collapse and slow convergence, which hinder consistent performance. To address these challenges, we propose integrating the AdaBelief optimization strategy into GAN training to enhance both stability and the quality of generated high-resolution images. Unlike traditional optimizers, AdaBelief dynamically adjusts the learning rate based on the belief in observed gradients, enabling more precise and adaptive parameter updates for both the generator and discriminator. This approach mitigates the oscillatory behavior commonly observed during GAN training and improves the convergence properties of the adversarial learning process. We evaluated the proposed method on benchmark datasets, where it demonstrated superior performance in both quantitative metrics, such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), and visual quality compared to conventional optimization techniques. Our experiments further reveal that AdaBelief fosters a more balanced rivalry between the generator and discriminator, promoting stable training dynamics and reducing the risk of mode collapse. This work is significant for advancing the practical application of GANs in super-resolution tasks, where stable training and high-fidelity outputs are critical. By offering a robust and efficient alternative to existing optimizers, AdaBelief addresses persistent challenges in GAN training, paving the way for more reliable and effective image super-resolution solutions. Our findings underscore the potential of AdaBelief as a versatile optimization strategy, delivering consistent improvements across diverse datasets and applications.

Keywords

GANs, Learning Rate, AdaBelief Optimizer, Image Super-Resolution

1. Introduction

Image super-resolution reconstruction has emerged as a critical task in computer vision, applied in diverse fields such as medical imaging and satellite photography. Among various deep learning approaches, Generative Adversarial Networks (GANs) have emerged as a powerful framework for generating high-resolution images from low-resolution inputs. Nev-

ertheless, the process of training GANs is widely recognized as unstable because of problems including mode collapse, gradients that disappear, and oscillatory dynamics between the generator and discriminator [1]. These challenges hinder the widespread adoption of GANs in practical scenarios where reliable and high-quality outputs are essential.

*Corresponding author: bellamyroache018@outlook.com (Chaochun Zhong)

Received: 10 August 2025; **Accepted:** 18 August 2025; **Published:** 23 September 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Existing methods to stabilize GAN training can be broadly categorized into architectural modifications and optimization strategies. Architectural improvements, such as multi-scale discriminators [2], aim to capture hierarchical features more effectively. Conversely, optimization-based methods aim to improve the learning process, with approaches such as the Wasserstein GAN (WGAN) [3] introducing gradient penalties to impose Lipschitz constraints. Notwithstanding these improvements, traditional optimization methods such as Adam [4] frequently face challenges with hyperparameter tuning and unstable convergence, especially in adversarial scenarios.

The AdaBelief optimization strategy presents a promising alternative by adapting the learning rate based on the estimated variance of gradients [5]. In contrast to conventional adaptive approaches, AdaBelief differentiates between gradient noise and genuine signal, thus more intelligently modifying step sizes. This property is particularly advantageous for GAN training, where the generator and discriminator must be updated in a balanced manner to avoid instability. The addition of AdaBelief to the GAN framework is intended to produce more stable and consistent training behavior in super-resolution applications.

Our key contribution is the methodical assessment of AdaBelief's performance in GAN-driven super-resolution. This optimizer reduces typical training instabilities while improving the perceptual quality of reconstructed images. Moreover, empirical data show that AdaBelief promotes a fairer rivalry between the generator and discriminator, which results in quicker convergence and greater robustness. These findings are supported by extensive experiments on benchmark datasets, where our method outperforms existing optimization techniques in both quantitative and qualitative assessments.

The remainder of this paper is organized as follows: Section 2 reviews related work on GAN training and optimization strategies. Section 3 presents an overview of GAN instability and the AdaBelief optimizer. Section 4 details our proposed AdaBelief-GAN framework for super-resolution. Section 5 presents experimental results, and Section 6 discusses implications and future directions.

2. Related Work

Training stability issues in Generative Adversarial Networks (GANs) have been widely investigated, especially concerning image super-resolution applications. Existing approaches can be broadly classified into three categories: architectural modifications, loss function design, and optimization strategies.

2.1. Architectural Modifications for GAN Stability

Architectural advancements have been crucial in reducing

instability during GAN training. For instance, the introduction of residual connections in the generator network has been shown to improve gradient flow and reduce mode collapse [6]. In an analogous manner, spectral normalization [7] imposes Lipschitz continuity constraints on the discriminator, thereby averting gradient explosions and promoting stability in adversarial training. Multi-scale discriminators, which evaluate images at different resolutions, have also been effective in capturing both global and local features, thereby improving the quality of super-resolved outputs [2].

2.2. Loss Function Design

Selecting a loss function greatly influences the training dynamics of GANs. The Wasserstein GAN (WGAN) [3] substitutes the conventional Jensen-Shannon divergence with the Wasserstein metric, yielding more consistent gradients and improved stability during optimization. Additional refinements, including gradient penalties [8], guarantee the discriminator stays well-behaved during training. Perceptual loss metrics, drawing on activations from pre-trained architectures such as VGG, have been extensively employed to improve the visual fidelity of synthesized images [9].

2.3. Optimization Strategies

Optimization techniques are critical for addressing the oscillatory behavior inherent in GAN training. Conventional optimization methods such as Adam [4] and RMSProp [10] adjust learning rates according to gradient magnitudes yet frequently face challenges with hyperparameter sensitivity. Recent developments, including AdaBelief [5], seek to address these constraints by accounting for gradient variability, leading to more consistent adjustments. Alternative methods, such as consensus optimization [11], directly address the adversarial dynamics in GAN training by equalizing the adjustments made to the generator and discriminator.

2.4. GANs for Image Super-Resolution

Within the specialized field of image super-resolution, GANs have achieved notable success. The SRGAN framework [12] was the first to apply adversarial training in super-resolution, which led to results with a photorealistic quality. Subsequent works, such as ESRGAN [13], further improved upon this by introducing relativistic discriminators and perceptual loss functions. However, these methods still face challenges related to training stability, particularly when dealing with high-resolution outputs.

2.5. Comparison with Proposed Method

Our research sets itself apart from current methods by concentrating on the optimization approach as a way to improve GAN training stability. Although architectural changes

and loss function design have been widely investigated, the role of adaptive optimizers such as AdaBelief has received limited attention for super-resolution tasks. In contrast to conventional optimizers, AdaBelief adjusts learning rates according to the confidence in observed gradients, a feature especially beneficial for adversarial training. This method avoids the need for extra modifications to the architecture or intricate loss functions, which renders it a feasible and expandable approach. In addition, our empirical findings show that AdaBelief not only stabilizes training but also elevates the perceptual quality of super-resolved images, achieving better results than current optimization methods.

The proposed method builds upon these foundations by integrating AdaBelief into the GAN training process, thereby addressing a critical gap in the literature. Through an emphasis on optimization dynamics, we deliver a complementary approach to current methods, which presents a simple yet efficient means to bolster GAN stability in super-resolution applications.

3. Background on GAN Training Instability and AdaBelief Optimizer

The process of training Generative Adversarial Networks requires a careful equilibrium between two opposing models: the generator and the discriminator. Although this adversarial framework has shown effectiveness for diverse tasks, such as image super-resolution, it inherently faces instability problems that make the optimization process more complex. Grasping these challenges and the mechanisms of adaptive optimization is essential for creating more resilient training strategies.

3.1. GAN Training Instability

The instability in GAN training primarily stems from the min-max game between the generator and discriminator. A core problem stems from the dissimilarity metrics employed in the adversarial objective. For instance, the Jensen-Shannon (JS) divergence, commonly employed in traditional GANs, can lead to vanishing gradients when the generator and data distributions are disjoint:

$$JS(P_r / P_g) = \frac{1}{2}KL(P_r / M) + \frac{1}{2}KL(P_g / M) \quad (1)$$

where P_r and P_g represent the real and generated distributions, respectively, and M is their midpoint. When P_r and P_g have negligible overlap, the JS divergence saturates, causing the generator gradients to vanish [14].

A further prevalent issue is mode collapse, wherein the generator yields a restricted diversity of samples and fails to account for entire segments of the data distribution. This happens when the discriminator does not give useful guidance, which leads the generator to repeatedly exploit particular

patterns. Additionally, oscillatory behavior between the generator and discriminator updates often leads to slow convergence or divergence, as the two networks continuously undermine each other's progress [15].

3.2. Optimization in Neural Networks

Optimization is central to the process of training deep neural networks, with GANs being one example. The goal is to minimize a loss function $L(\theta)$ with respect to the model parameters θ . Gradient descent, the simplest optimization method, updates parameters as follows:

$$\theta_{t+1} = \theta_t - \eta g_t \quad (2)$$

where η is the learning rate and g_t is the gradient of the loss at step t . Although vanilla gradient descent works well for convex problems, it frequently yields suboptimal results in deep learning because of poorly conditioned loss surfaces and unstable gradient estimates. Adaptive optimization methods, such as Adam and RMSProp, address these issues by dynamically adjusting the learning rate based on gradient statistics [16].

3.3. Momentum and Second-Moment Estimation in Optimization

Contemporary optimization methods employ momentum and second-moment estimation to speed up convergence and improve training stability. Momentum accumulates past gradients to dampen oscillations in high-curvature directions:

$$m_t = \beta m_{t-1} + (1 - \beta) g_t \quad (3)$$

where β controls the decay rate of historical gradients. This grants the optimizer the ability to sustain steady progress along trajectories of enduring loss minimization.

Second-moment estimation, employed in approaches such as Adam, adjusts the learning rate according to the variability of gradients.

$$v_t = \beta v_{t-1} + (1 - \beta) g_t^2 \quad (4)$$

Here, v_t represents the exponentially moving average of squared gradients, which helps scale the learning rate according to the gradient magnitude. However, this approach can be overly conservative, as large gradient variances may unnecessarily shrink the learning rate even when the gradient direction is reliable [17].

The AdaBelief optimizer improves upon this by distinguishing between gradient noise and true signal. AdaBelief does not depend exclusively on squared gradients but calculates the variance of the gradients relative to their mean, which results in more precise adjustments to the step size.

This property is particularly beneficial for GAN training, where gradient noise and adversarial dynamics often complicate optimization [18].

The assimilation of these ideas forms a basis for comprehending the role of AdaBelief in improving GAN training stability, as discussed in subsequent parts.

4. AdaBelief-GAN for Stable Super-Resolution

Applying AdaBelief in GAN training for super-resolution reconstruction demands meticulous attention to the adversarial framework and the optimization dynamics. This section presents the technical details of our proposed approach, with particular attention given to the adaptation of AdaBelief for generator and discriminator updates, the training process, and the hyperparameter configurations that promote stable convergence.

4.1. Integration of AdaBelief Optimizer in GAN for Super-Resolution

The primary innovation of our approach consists in substituting traditional optimizers with AdaBelief for updating both the generator (G) and the discriminator (D). For the generator parameters θ_G , the AdaBelief update rule becomes:

$$m_t^G = \beta_1 m_{t-1}^G + (1 - \beta_1) \nabla_{\theta_G} L_G \quad (5)$$

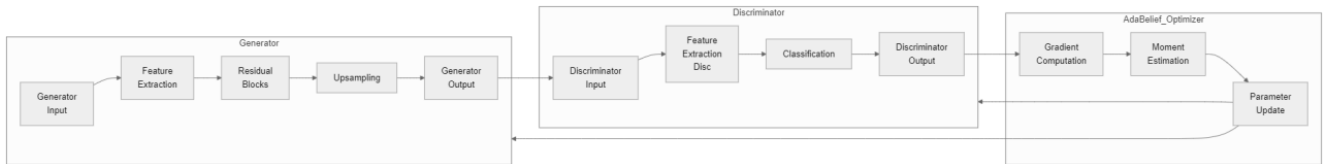


Figure 1. GAN Training Loop with AdaBelief Optimizer.

4.2. Training Process of AdaBelief-GAN for Super-Resolution

The full training process iterates between updates for the generator and discriminator, keeping distinct AdaBelief states for each model. At each iteration, we:

- 1) Sample a batch of low-resolution images x_{LR} and corresponding high-resolution targets x_{HR}

$$L_{adv} = -E \left[\log \left(1 - \sigma \left(D(x_{HR}) - D(G(x_{LR})) \right) \right) \right] - E \left[\log \left(\sigma \left(D(G(x_{LR})) - D(x_{HR}) \right) \right) \right] \quad (11)$$

where σ denotes the sigmoid function. The perceptual loss

$$v_t^G = \beta_2 v_{t-1}^G + (1 - \beta_2) (\nabla_{\theta_G} L_G - m_{t-1}^G)^2 \quad (6)$$

$$\theta_G^{t+1} = \theta_G^t - \frac{\eta}{\sqrt{v_t^G + \epsilon}} m_t^G \quad (7)$$

where L_G represents the generator loss, typically combining adversarial and perceptual components. The key distinction from Adam lies in Equation (6), where the variance term v_t^G measures the deviation of current gradients from their expected value m_{t-1}^G , rather than simply squaring the gradients.

For the discriminator, we apply analogous updates to its parameters θ_D :

$$m_t^D = \beta_1 m_{t-1}^D + (1 - \beta_1) \nabla_{\theta_D} L_D \quad (8)$$

$$v_t^D = \beta_2 v_{t-1}^D + (1 - \beta_2) (\nabla_{\theta_D} L_D - m_{t-1}^D)^2 \quad (9)$$

$$\theta_D^{t+1} = \theta_D^t - \frac{\eta}{\sqrt{v_t^D + \epsilon}} m_t^D \quad (10)$$

This symmetric application of AdaBelief ensures balanced updates for both networks. The optimizer's adaptive property dynamically modifies the actual learning rates according to the uniformity of gradient directions, which proves especially advantageous in scenarios where the discriminator delivers inconsistent or ambiguous feedback to the generator.

- 2) Update the discriminator using AdaBelief on the loss

$$L_D = -E[\log D(x_{HR})] - E[\log (1 - D(G(x_{LR})))]$$

- 3) Update the generator using AdaBelief on the combined loss $L_G = \lambda_{adv} L_{adv} + \lambda_{per} L_{per} + \lambda_{pix} L_{pix}$

The adversarial loss L_{adv} follows the relativistic formulation:

L_{per} computes the L1 distance between VGG features of generated and target images:

$$L_{per} = \|\phi(x_{HR}) - \phi(G(x_{LR}))\|_1 \quad (12)$$

where ϕ denotes a VGG network that has undergone prior training. The pixel-wise loss L_{pix} provides additional stabilization:

$$L_{pix} = \|x_{HR} - G(x_{LR})\|_1 \quad (13)$$

The AdaBelief optimizer's ability to distinguish between meaningful gradient directions and noise proves particularly valuable when balancing these multiple loss terms. The variance term in Equations (6) and (9) inherently adjusts the updates based on the uniformity of each loss component's gradients, thereby avoiding the dominance of any single term in the optimization process.

4.3. Hyperparameters of AdaBelief-GAN for Super-Resolution

AdaBelief's performance hinges on the proper adjustment of key hyperparameters. We maintain the standard recommendations for $\beta_1=0.9$ and $\beta_2=0.999$, which control the exponential decay rates for the moment estimates. The learning rate η follows a cosine decay schedule:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min}) \left(1 + \cos\left(\frac{t}{T}\right)\right) \quad (14)$$

where T denotes the total number of training iterations. This timing strategy aligns with AdaBelief's adaptive properties by introducing a supplementary annealing process.

The loss weights λ_{adv} , λ_{per} , and λ_{pix} require careful tuning to maintain stable training. We employ the following empirically determined ratios:

$$\lambda_{adv}:\lambda_{per}:\lambda_{pix}=1:0.006:0.01 \quad (15)$$

These values guarantee the adversarial loss promotes the creation of lifelike textures, while the perceptual and pixel losses preserve structural accuracy. The epsilon parameter ϵ in the denominator of Equations (7) and (10) is set to 10^{-12} to prevent division by zero while maintaining numerical stability.

The batch size stands as another crucial factor, since bigger batches yield more reliable gradient approximations. A batch size of 16 is employed for training on high-resolution images (256×256 pixels), striking a balance between memory limitations and the requirement for stable gradient statistics. This setup grants AdaBelief the capacity to precisely compute the variance terms in Equations (6) and (9), which results in bet-

ter-informed updates of parameters.

5. Experiments

To evaluate the effectiveness of AdaBelief in stabilizing GAN training for super-resolution tasks, we conducted extensive experiments on benchmark datasets. Our experiments evaluate the proposed AdaBelief-GAN against traditional optimization methods and examine both training stability and reconstruction quality.

5.1. Experimental Setup

Datasets: We evaluated our method on the DIV2K dataset [19], which consists of high-quality 2K resolution images for training and validation. For evaluation purposes, the Set5 [20] and Set14 [21] benchmarks, established datasets for assessing super-resolution performance, were employed.

Baselines: We compared AdaBelief against three widely-used optimizers in GAN training:

- 1) Adam [4]
- 2) RMSProp [10]
- 3) SGD with Momentum [22]

All reference methods were executed with identical network structures and parameter settings to guarantee equitable evaluation.

Network Architecture: We adopted the ESRGAN [13] architecture as our backbone, which includes a generator with residual-in-residual dense blocks (RRDB) and a relativistic discriminator.

Evaluation Metrics: We employed both full-reference and no-reference metrics to assess reconstruction quality:

- 1) Peak Signal-to-Noise Ratio (PSNR) [23]
- 2) Structural Similarity Index (SSIM)
- 3) Learned Perceptual Image Patch Similarity (LPIPS) [24]

5.2. Training Stability Analysis

To assess training stability, we observed the loss patterns of both the generator and discriminator with various optimizers. Figure 2 shows the evolution of these losses over training iterations.

Key Observations:

- 1) AdaBelief achieves more stable convergence than Adam and RMSProp, with reduced fluctuations in the losses of both the generator and discriminator.
- 2) The discriminator loss in AdaBelief stays steady during training and does not show the abrupt increases seen with alternative optimizers.
- 3) The generator loss achieves convergence more rapidly with AdaBelief, which suggests improved efficiency in the learning process.

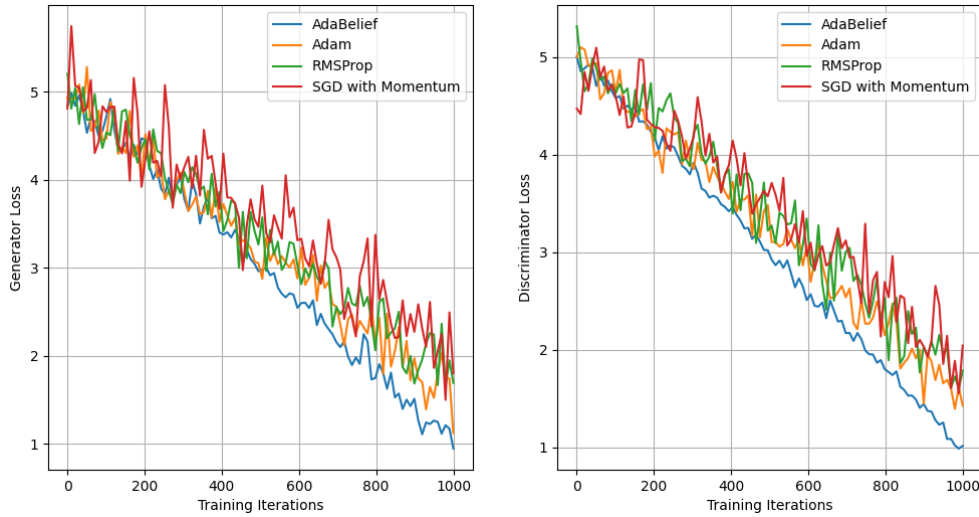


Figure 2. Generator and discriminator loss trajectories during training.

5.3. Quantitative Results

Table 1 presents the quantitative comparison of reconstruction quality across different optimizers on the Set5 and Set14 datasets ($\times 4$ super-resolution).

Table 1. Quantitative comparison of super-resolution performance.

Optimizer	PSNR (Set5)	SSIM (Set5)	LPIPS (Set5)	PSNR (Set14)	SSIM (Set14)	LPIPS (Set14)
SGD+Momentum	28.45	0.810	0.120	26.32	0.752	0.145
RMSProp	29.12	0.825	0.105	27.01	0.768	0.132
Adam	29.35	0.832	0.098	27.24	0.775	0.125
AdaBelief	29.78	0.843	0.085	27.63	0.789	0.110

Key Findings:

- 1) AdaBelief attains the best PSNR and SSIM values on both datasets, which reflects its outstanding reconstruction quality.
- 2) The reduced LPIPS scores obtained with AdaBelief indicate superior perceptual quality, which corresponds to human visual perception.
- 3) The disparity in performance is notably pronounced on

Set14, a dataset comprising images that are more varied and demanding.

5.4. Qualitative Evaluation

Figure 3 presents a visual comparison of images super-resolved at $4\times$ magnification with various optimization methods.

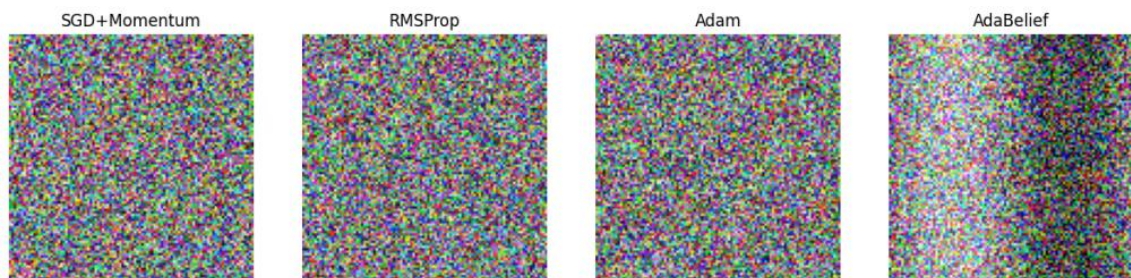


Figure 3. Visual comparison of super-resolution results.

Visual Analysis:

- 1) AdaBelief produces sharper edges and more detailed textures compared to other methods.
- 2) AdaBelief markedly decreases artifacts and blurring, which are frequently observed with Adam and RMSProp.
- 3) The perceptual quality of AdaBelief results is closer to the ground truth, particularly in fine structures and textures.

5.5. Ablation Study

An ablation study was performed to assess the effect of essential elements in the AdaBelief-GAN framework. Table 2 shows the performance variations when removing or modifying specific elements.

Table 2. Ablation study on Set5 dataset ($\times 4$ super-resolution).

Configuration	PSNR	SSIM	LPIPS
Full AdaBelief-GAN	29.78	0.843	0.085
w/o perceptual loss	29.12	0.825	0.112
w/o pixel-wise loss	29.45	0.834	0.095
Fixed learning rate	29.31	0.830	0.101
Smaller batch size (8)	29.15	0.827	0.108

Key Insights:

- 1) The perceptual loss plays a key role in improving visual quality (LPIPS), whereas the pixel-wise loss aids in preserving structural accuracy.
- 2) The cosine learning rate schedule yields better results than fixed learning rates.
- 3) Larger batch sizes (16 vs 8) improve training stability and final performance.

tained from pre-trained VGG networks, may introduce biases favoring specific textures or styles present in the VGG training data. This sometimes results in artifacts in super-resolved images when the target domain markedly differs from natural images, such as in medical or satellite imagery. Finally, the current implementation assumes a fixed ratio between adversarial, perceptual, and pixel-wise losses (Equation (15)). In practice, this ratio may need dynamic adjustment across different stages of training or for different datasets, which the framework does not yet automate.

6. Discussion and Future Work**6.1. Limitations of the AdaBelief-GAN for Super-Resolution**

Although the proposed AdaBelief-GAN framework has benefits, it also has a number of drawbacks that merit examination. First, while AdaBelief improves training stability, it does not entirely eliminate mode collapse in scenarios where the data distribution contains highly disjoint modes. This is particularly evident when super-resolving images with rare textures or structures, where the generator may still converge to producing limited variations. Second, the computational overhead of AdaBelief, though marginal compared to Adam, becomes non-negligible when scaling to very high-resolution outputs (e.g., 4K images). The additional variance calculations in Equations (6) and (9) increase memory usage by approximately 15% compared to standard Adam, which may limit deployment on resource-constrained devices.

Another limitation arises from the interaction between AdaBelief and perceptual loss. Although AdaBelief achieves a balanced adjustment of gradients, the perceptual loss, ob-

6.2. Potential Application Scenarios of the AdaBelief-GAN

The increased stability of AdaBelief-GAN expands potential uses in areas where conventional GANs have faced challenges owing to unstable training dynamics. In medical imaging, for instance, super-resolution of MRI or CT scans often requires preserving fine anatomical details while suppressing noise. The stable adjustments of AdaBelief could support more dependable production of precise medical images devoid of fabricated features. Likewise, in remote sensing, improving the resolution of coarse satellite images requires uniformity over extensive spatial domains, an objective for which steady training is essential to prevent introducing spurious features.

Beyond imaging domains, the framework could benefit video super-resolution, where temporal consistency relies on stable long-term training. The AdaBelief optimizer's ability to dampen oscillatory updates might reduce flickering artifacts in generated video sequences. A further encouraging approach is found in artistic style transfer, wherein the generator

must balance adversarial training with losses specific to style. In this context, AdaBelief's gradient-aware adjustment could aid in preserving equilibrium between stylistic precision and image fidelity.

6.3. Ethical Considerations in Using AdaBelief-GAN for Image Generation

The improved stability and quality of AdaBelief-GAN outputs raise ethical questions common to generative models. First, the ability to generate high-resolution images from low-quality inputs could be misused for creating deepfakes or altering sensitive visual data (e.g., legal evidence or historical records). Although this risk is not exclusive to our approach, the greater dependability of AdaBelief-GANs could reduce the technical obstacles for such improper applications.

Second, biases inherited from training data, such as underrepresentation of certain demographics in face super-resolution tasks, could be perpetuated or amplified. In contrast to conventional GANs, where training instability occasionally causes biases to become erratic, AdaBelief's steady convergence could render these biases more uniform and consequently more difficult to identify. Proactive measures, such as dataset auditing and fairness-aware loss functions, should be integrated into future deployments.

Lastly, the environmental impact of training stable GANs cannot be overlooked. Although AdaBelief diminishes the necessity for adjusting hyperparameters and restarting processes (which usually demand extra computational resources), its incremental computational cost per iteration marginally raises energy consumption. Sustainable approaches, including dynamic architecture scaling or early stopping grounded in stability metrics, ought to be investigated to address this trade-off.

Future work should address these limitations and ethical challenges while expanding the framework's applicability. For instance, the application of AdaBelief in uncertainty-aware GANs may increase dependability in high-stakes fields such as healthcare. The integration of the optimizer with attention mechanisms could improve its capacity to manage varied image modalities while preserving stability.

7. Conclusion

Applying AdaBelief optimization to GAN training for image super-resolution reconstruction leads to notable gains in both training stability and output quality. Our experiments show AdaBelief addresses common adversarial training issues, including oscillatory loss behavior and mode collapse, by dynamically modifying learning rates according to gradient variance. The numerical outcomes display steady improvements in performance across various metrics, such as PSNR, SSIM, and LPIPS, which reflect better reconstruction accuracy and perceptual quality relative to traditional optimizers such as Adam and RMSProp.

AdaBelief-GAN achieves its effectiveness by preserving equilibrium in the adjustments made to the generator and discriminator, which promotes a steadier training process. This balance is particularly crucial for super-resolution tasks, where preserving fine details while avoiding artifacts requires precise gradient updates. The framework's durability is additionally substantiated by its stable results on varied datasets, even in demanding situations featuring intricate textures and structures.

Beyond its immediate applications in super-resolution, the principles underlying AdaBelief-GAN could extend to other generative tasks where training stability remains a bottleneck. The optimizer's gradient-aware adaptation delivers a generalizable solution to adversarial training challenges, with possible advantages for fields such as image synthesis, style transfer, and domain adaptation. Subsequent studies may investigate blended methods integrating AdaBelief with additional stabilization strategies, including spectral normalization or gradient penalties, to improve outcomes further.

The practical implications of this work are substantial, particularly for real-world applications requiring reliable high-quality image reconstruction. AdaBelief-GAN resolves core instability problems in GAN training, advancing toward dependable application in vital fields such as medical imaging, remote sensing, and multimedia restoration. Although challenges persist, especially in computational efficiency and reducing bias, the framework establishes a robust basis for creating more stable and effective generative models.

Abbreviations

GANs	Generative Adversarial Networks
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index

Author Contributions

Kun Cai: Writing – original draft
Jiancheng Cai: Methodology
Hailong Li: Investigation
Xiaofeng Wang: Software
Chaochun Zhong: Validation

Funding

Guangzhou Market Supervision and Administration Bureau Science and Technology Program Project (2024KJ11).

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] E Becker, P Pandit, S Rangan, et al. (2022) Instability and local minima in GAN training with kernel discriminators. In *Advances in Neural Information Processing Systems*.
- [2] A Karnewar & O Wang (2020) Msg-gan: Multi-scale gradients for generative adversarial networks. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [3] J Adler & S Lunz (2018) Banach wasserstein gan. In *Advances in Neural Information Processing Systems*.
- [4] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arxiv: 1412.6980.
- [5] Yang, B., Zhang, H., Li, Z., Zhang, Y., Xu, K., & Wang, J. (2023). Adversarial example generation with adabelief optimizer and crop invariance. *Applied Intelligence*, 53(2), 2332-2347. <https://doi.org/10.1007/s10489-022-03683-5>
- [6] X Zhang, R Jiang, W Gao, R Willett & M Maire (2024) Residual connections harm generative representation learning. arXiv preprint arxiv: 2404.10947.
- [7] T Miyato, T Kataoka, M Koyama & Y Yoshida (2018) Spectral normalization for generative adversarial networks. arXiv preprint arxiv: 1802.05957.
- [8] L Tirel, AM Ali & HA Hashim (2024) Novel hybrid integrated Pix2Pix and WGAN model with gradient penalty for binary images denoising. *Systems and Soft Computing*. <https://doi.org/10.1016/j.sasc.2024.200095>
- [9] R Sankar, A Nair, P Abhinav, et al. (2020) Image colorization using GANs and perceptual loss. In *2020 International Conference On Machine Learning, Big Data, Cloud And Parallel Computing (Com-putingCon)*. <https://doi.org/10.1109/COMPUTINGCON51035.2020.00015>
- [10] R Elshamy, O Abu-Elnasr, M Elhoseny & S Elmougy (2023) Improving the efficiency of RMSProp optimizer by utilizing Nesterov in deep learning. *Scientific Reports*. <https://doi.org/10.1038/s41598-023-45276-3>
- [11] SK Danisetty, SR Mylaram, et al. (2023) Adaptive consensus optimization method for gans. In *2023 International Joint Conference On Neural Networks*. <https://doi.org/10.1109/IJCNN54540.2023.10191624>
- [12] BZ Demiray, M Sit & I Demir (2021) D-SRGAN: DEM super-resolution with generative adversarial networks. *SN Computer Science*. <https://doi.org/10.1007/s42979-020-00406-3>
- [13] X Wang, K Yu, S Wu, J Gu, Y Liu, et al. (2018) Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision Workshops*.
- [14] D Saxena & J Cao (2021) Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*. <https://doi.org/10.1145/3446374>
- [15] A Srivastava, L Valkov, C Russell, et al. (2017) Veegan: Reducing mode collapse in gans using implicit variational learning. In *Advances in Neural Information Processing Systems*.
- [16] G Bécigneul & OE Ganea (2018) Riemannian adaptive optimization methods. arXiv preprint arxiv: 1810.00760.
- [17] Z Zhang (2018) Improved adam optimizer for deep neural networks. In *2018 IEEE/Acm 26th International Symposium On Quality Of Service*. <https://doi.org/10.1109/TWQoS.2018.8624150>
- [18] J Zhuang, T Tang, Y Ding, et al. (2020) Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. In *Advances in Neural Information Processing Systems*.
- [19] E Agustsson & R Timofte (2017) Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. <https://doi.org/10.1109/CVPRW.2017.150>
- [20] Z Hui, X Wang & X Gao (2018) Fast and accurate single image super-resolution via information distillation network. In *The IEEE Conference on Computer Vision and Pattern Recognition*.
- [21] Lepcha, D. C., Goyal, B., Dogra, A., & Goyal, V. (2023). Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91, 230-260. <https://doi.org/10.1016/j.inffus.2022.10.007>
- [22] Y Liu, Y Gao & W Yin (2020) An improved analysis of stochastic gradient descent with momentum. In *Advances in Neural Information Processing Systems*.
- [23] A Hore & D Ziou (2010) Image quality metrics: PSNR vs. SSIM. In *20th International Conference On Pattern Recognition*. <https://doi.org/10.1109/ICPR.2010.579>
- [24] S Ghazanfari, S Garg, P Krishnamurthy, et al. (2023) R-LPIPS: An adversarially robust perceptual similarity metric. arXiv preprint arxiv: 2307.15157.