Artificial Intelligence Research and Applications (AIra)
https://www.ijaira.org/ Online ISSN: 3108-3889
E. Demir et al., Vol. 2, No. 2, June, 2026 Research Article

Loss-Level Stabilization in Deep Image Denoising via Fractional-Inspired Memory-Weighted Regularization

Elif Demir *, Yusuf Zeren *,**, Alpaslan Demirci ***‡

* Department of Mathematics, Faculty of Science and Literature, Yildiz Technical University, Istanbul, Türkiye.
** Department of Statistics, Faculty of Humanities and Social Sciences, Istanbul Ticaret University, Istanbul, Türkiye.
*** Department of Electrical Engineering, Faculty of Electrical and Electronics Engineering, Yildiz Technical University, Istanbul, Türkiye.

(elifdemir1317@gmail.com, yzeren@yildiz.edu.tr, ademirci@yildiz.edu.tr)

‡ Corresponding Author: Alpaslan Demirci, Department of Electrical Engineering, Faculty of Electrical and Electronics Engineering, Yildiz Technical University, Istanbul, Türkiye, ademirci@yildiz.edu.tr.

Received: 18.02.2026, Revised: 10.04.2026, Accepted: 24.04.2026

Abstract- This study investigates the role of memory-weighted (fractional-inspired) regularization in stabilizing training dynamics for deep image denoising. Unlike conventional loss functions that rely solely on instantaneous residuals, the proposed approach introduces a loss-level aggregation mechanism in which past residuals are incorporated through a weighted memory structure. Specifically, a memory-weighted regularization term is combined with a base loss function, where historical residuals are aggregated using a power-law decay weighting scheme. This formulation does not correspond to a strict fractional derivative operator, but is inspired by the long-memory characteristics of fractional calculus. The method is evaluated under controlled experimental settings across multiple noise intensities and noise types, including Gaussian and impulsive noise. All experiments are conducted using identical network architectures and training configurations to ensure fair comparison. The results show that, while improvements in reconstruction metrics such as PSNR and SSIM are modest, the proposed formulation consistently reduces late-stage fluctuation and yields smoother convergence behaviour. In particular, the reduction in PSNR standard deviation indicates improved optimization stability under high-noise conditions. These findings suggest that memory-aware loss design provides a complementary perspective for improving convergence stability in deep learning-based image denoising.

Keywords: Image denoising, memory-weighted regularization, fractional-inspired loss, training stability, loss function design.

1. Introduction

Image denoising is a fundamental problem in computer vision and image processing, serving both as a basic pre-processing step and as a critical component that directly affects the performance of many high-level tasks such as recognition, segmentation, and reconstruction [1- 3]. Real-world images inevitably contain noise due to various factors, including sensor limitations, low-light conditions, compression artifacts, and transmission errors [4, 5]. This noise not only causes pixel- level degradation but also alters the statistics of structural components such as edges, textures, and fine details, thereby complicating the training and generalization behaviour of learning-based models [6]. Particularly in deep learning-based approaches, increasing

Cite this article as: E. Demir, Y. Zeren, and A. Demirci, "Loss-Level Stabilization in Deep Image Denoising via Fractional-Inspired Memory-Weighted Regularization," *Artificial Intelligence Research and Applications (AIra)*, Vol. 2, No. 2, pp. 131–139, June, 2026, https://doi.org/10.20508/5j3zde56.

noise levels frequently induce optimization instability, convergence degradation, and visually noticeable artifacts [7-9]. Consequently, recent literature emphasizes that denoising performance depends not only on network architecture but also on the design of the loss function that governs the learning process [10, 11].

In deep learning-based denoising models, the most commonly used loss functions are based on squared error ($L2$) and absolute error ($L1$) formulations [12, 13]. These losses are widely adopted due to their mathematical simplicity and stable large-scale training behaviour [14]. However, especially in the case of $L2$-based losses, the quadratic amplification of large residuals can cause rare but extreme pixel deviations—such as those produced by impulsive noise—to dominate the optimization direction [15]. In practice, this behaviour often manifests as oscillatory training curves, delayed convergence, and local distortions in reconstructed outputs [16]. Robust loss functions such as Huber, Cauchy, and related variants have therefore been proposed to mitigate outlier sensitivity [17]. By limiting gradient amplification for large residuals, robust losses offer improved stability under challenging noise conditions. Nevertheless, when deployed with fixed parameter settings, their adaptive capacity across varying noise intensities and distributions may remain limited [18]. These observations highlight a key issue: beyond the analytical form of a loss function, its gradient-generation characteristics throughout training critically determine the stability and smoothness of the optimization trajectory [9, 17].

Fractional calculus provides a mathematically grounded perspective that is intrinsically associated with memory effects and long-range dependencies, and has demonstrated effectiveness in signal processing, control systems, and dynamical modelling [19- 21]. Unlike classical integer-order derivatives, fractional derivatives incorporate the accumulated influence of past states, offering a memory-aware description of system behaviour. Such memory-driven formulations are frequently linked to smoother and more balanced responses under irregular or abruptly changing conditions [21- 23]. In image processing, fractional operators have been successfully employed in filtering, edge enhancement, and texture analysis tasks [24, 25]. However, the use of fractional intuition as a regulatory principle for shaping the learning-time behaviour of loss functions in deep models remains relatively unexplored [21]. This gap motivates a systematic examination of how fractional memory concepts can complement robust loss formulations from a training-dynamics perspective.

Building on this motivation, we study memory-weighted (fractional-inspired) regularization as a loss-level mechanism for stabilizing learning dynamics in deep image denoising. The central design choice is to intervene only in the objective function: a memory-weighted term is composed with a base loss so that residual information is aggregated in a history-aware manner during training, while the network architecture, optimizer, and training schedule remain fixed. This controlled formulation enables an explicit analysis of (i) outlier sensitivity and (ii) convergence oscillations under identical computational conditions. Beyond reporting final PSNR/SSIM values, we quantify late-stage stability using fluctuation statistics computed over the final training window, thereby connecting reconstruction quality with optimization behaviour. Across challenging noise levels, the memory-weighted loss construction produces smoother convergence trajectories and reduced fluctuation intensity, without sacrificing reconstruction fidelity.

The contributions of this work can be summarized as follows:

- A conceptual evaluation framework is introduced that integrates fractional memory intuition with robust loss design from the perspective of learning stability.
- The influence of loss-function behaviour on outlier sensitivity and convergence oscillations is systematically examined, providing intuitive insights into how memory-weighted regularization can mitigate instability.
- Through controlled experimental analysis, it is shown that incorporating memory-weighted regularization into loss structures promotes more consistent convergence patterns and reduced fluctuation under high-noise conditions.

While a wide range of loss functions have been proposed for image denoising, including robust losses (e.g., $L1$, LogCosh, Cauchy) and adaptive loss formulations, these approaches primarily operate on instantaneous residuals and do not explicitly model temporal error persistence across training iterations. In contrast, the proposed approach introduces a memory-weighted aggregation mechanism that incorporates historical residual information into the loss formulation. Unlike classical smoothing or momentum-based techniques, this formulation operates directly at the loss level and systematically reweights past residual contributions. This distinction positions the proposed method as a loss-level memory modelling strategy rather than a modification of the optimization process. As a result, the method provides a complementary perspective to existing robust and adaptive loss designs, specifically targeting convergence stability rather than solely robustness to outliers. In this context, the present study focuses on improving optimization stability in deep image denoising, which is a critical factor for the reliable deployment of learning-based image restoration systems. Rather than proposing a new architecture, the contribution lies in introducing a loss-level memory-weighted regularization mechanism that provides a complementary perspective on training dynamics and convergence behaviour.

The remainder of the paper is organized as follows. Section 2 reviews related work on robust loss functions and fractional calculus-based approaches in the context of image denoising. Section 3 introduces the fractional calculus perspective and the conceptual regularization view based on memory effects. Section 4 describes the proposed application framework and explains how the fractional regularization component is integrated into the training pipeline. Section 5 presents the problem formulation and the noise models considered. Section 6 details the experimental setup, fair comparison settings, and evaluation metrics. Section 7 discusses the quantitative and visual results from the perspective of learning stability. Finally, Section 8 provides concluding remarks and outlines directions for future research.

2. Related Work

Deep learning-based approaches to image denoising have achieved significant progress in recent years, and the resulting performance is closely tied to how the employed loss functions respond to the error signal during training [1- 3]. It has been widely reported that commonly used squared error ($L2$) and absolute error ($L1$) losses may destabilize the learning process under high noise levels and in scenarios with dense outlier pixel values [7- 9]. In particular, the disproportionate amplification of large residuals by the $L2$ loss can induce gradient oscillations and irregular convergence behaviour under impulsive noise conditions [15, 16]. These findings reinforce the view that loss design directly shapes optimization dynamics, rather than merely determining the final reconstruction metric.

Robust loss functions such as Huber, Cauchy, and related variants have therefore been proposed as alternatives that provide a more balanced response to outliers [17, 14]. By limiting the dominance of rare but large residuals in the update direction, these formulations enhance stability under challenging noise distributions and reduce the risk of gradient amplification caused by extreme deviations. Nevertheless, robust losses are commonly employed with fixed parameterizations and are predominantly evaluated in terms of final reconstruction accuracy, leaving their influence on convergence smoothness and stability under varying noise intensities and outlier densities comparatively less examined [18]. This observation motivates the development of loss constructions that regulate training dynamics at the level of residual aggregation, rather than relying solely on the pointwise analytical form of the objective. In this context, the broader question concerns how the structural behaviour of loss functions can be systematically shaped to further enhance training stability. Parallel to these developments, fractional calculus-based methods have been extensively utilized in signal processing, filtering, and control theory owing to their intrinsic capacity to model memory effects and long-range dependencies [20, 22]. Within image processing, fractional derivatives and integrals have been shown to produce smoother and more stable responses in tasks such as edge preservation, noise suppression, and texture analysis [6, 24-26]. More recently, fractional calculus has also been explored in broader machine learning contexts, including dynamical modelling and optimization-related formulations [19,21], thereby providing a conceptual bridge between memory-aware modelling principles and the regulation of learning dynamics. Despite these advances, fractional concepts are most commonly exploited at the signal or model level (e.g., filtering and priors) rather than as explicit regulators of learning-time loss behaviour in deep denoisers [21]. As a result, robust loss design and fractional memory modelling have largely evolved as parallel directions: robust objectives target outlier resistance, whereas fractional formulations emphasize history dependence and smooth responses. The present study connects these perspectives through an explicit loss-level composition, and evaluates how memory-weighted residual aggregation interacts with robust losses to improve convergence stability in denoising.

Existing approaches in the literature can be broadly categorized into robust loss functions, adaptive loss formulations, and optimization-level smoothing techniques. Robust loss functions aim to reduce sensitivity to outliers by modifying the penalty applied to large residuals, while adaptive losses dynamically adjust their shape during training. Optimization-level approaches, such as momentum or temporal smoothing, influence parameter updates but do not directly modify the structure of the loss function itself. The proposed method differs from these approaches by introducing a memory-weighted loss formulation that explicitly incorporates past residual information into the objective function. This design allows the method to capture temporal consistency in error patterns, which is not directly addressed by conventional robust or adaptive loss functions. Therefore, the proposed approach should be viewed not as a replacement for existing methods, but as a complementary framework that targets stability through structured memory integration at the loss level. This distinction ensures that the proposed approach addresses a different aspect of the problem, focusing on temporal consistency in residual aggregation rather than solely robustness or adaptivity.

3. Fractional-Inspired Memory Weighted Regularization: A Conceptual Perspective

Fractional calculus extends classical integer-order formulations by introducing operators whose behaviour reflects not only instantaneous information but also the accumulated influence of past states [19- 22]. This intrinsic memory effect is widely associated with smoother and more balanced responses in noisy, irregular, or abruptly varying environments. In the present study, fractional intuition is elevated from a signal-processing tool to a structural design principle for loss construction. Specifically, a memory-weighted residual aggregation mechanism is integrated into the objective function so that persistent error patterns across iterations are systematically emphasized during gradient formation. Through this formulation, the study establishes a direct connection between fractional memory concepts and measurable optimization behaviour, positioning loss-level memory modelling as a principled mechanism for regulating convergence stability in deep denoising.

Classical (integer-order) loss functions generate gradients that depend exclusively on the instantaneous magnitude of residual errors, such that the update direction at each iteration is determined primarily by the current deviation between prediction and ground truth, without accounting for the temporal evolution of error patterns. Under high-noise conditions—particularly in the presence of dense or impulsive outliers—this instantaneous sensitivity may cause disproportionate amplification of large residuals, leading to abrupt increases in gradient norms and irregular update directions [15- 17]. Such behaviour can manifest as oscillatory training dynamics, unstable convergence trajectories, and reduced consistency in parameter evolution. Beyond purely numerical instability, these fluctuations may disturb the balance between noise suppression and structural detail preservation, ultimately influencing reconstruction fidelity. In contrast, memory-weighted regularization introduces a fundamentally different aggregation principle by incorporating information about the persistence of residual patterns across successive iterations [19, 21]. Rather than reacting solely to instantaneous deviations, this mechanism redistributes emphasis toward residual components that exhibit sustained behaviour while attenuating the influence of isolated extreme errors. This redistribution does not merely scale gradient magnitudes; it reshapes the aggregation of error contributions at a structural level, yielding smoother update trajectories and more coherent optimization dynamics. Consequently, convergence becomes less susceptible to transient noise spikes, and overall training stability is enhanced without modifying the underlying network architecture or optimization algorithm.

In order to improve the clarity and reproducibility of the proposed approach, the memory-weighted regularization term is defined explicitly as follows. Let $r_t = \hat{y}_t - y_t$ denote the residual at iteration t, where $\hat{y}_t$ is the network output and y_t is the corresponding ground truth. Unlike classical loss formulations that depend only on the instantaneous residual, the proposed method aggregates residual information over a finite memory window. The memory-weighted regularization term is defined as a weighted sum of past residuals:

$$\mathcal{R}_{frac} = \sum_{k=0}^{M} w_k \phi(r_{t-k}) \tag{1}$$

where M denotes the memory size, $\mathrm{w_k}$ represents the weight assigned to the residual at lag k, and $\phi(.)$ is a pointwise penalty function (e.g., absolute value, squared error, or a robust loss such as LogCosh). In the experiments, $\phi(r)$ is instantiated as the LogCosh loss unless otherwise stated. In this study, the weights are defined using a decaying structure inspired by fractional-order memory behaviour, particularly through power-law decay characteristics: $\mathrm{w_k} = \frac{1}{(k+1)^{\alpha}}$ where $\alpha \in (0,1]$ controls the strength of the memory effect. This weighting follows a power-law decay structure, which is commonly associated with fractional-order memory behaviour. Smaller values of α assign more importance to long-term residuals, while larger values emphasize more recent errors. The total loss function is then formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{base} + \lambda \mathcal{R}_{frac} \tag{2}$$

where $\mathcal{L}_{base}$ denotes the base loss function (such as $L2, L1$, or LogCosh), and $\lambda > 0$ is a weighting parameter that balances instantaneous error minimization and memory-based regularization. In practice, λ is selected from a small range (e.g., $\lambda \in [0.01, 0.1]$) to ensure that the regularization term stabilizes training without dominating the base loss. From an implementation perspective, a residual buffer of size M is maintained during training. At each iteration, the current residual is appended to the buffer, and the oldest entry is removed when the buffer exceeds its capacity. The regularization term is computed by applying the weight sequence $\{w_k\}$ to the stored residual history. It is important to note that the proposed formulation is not a strict fractional derivative operator, but rather a fractional-inspired memory-weighted aggregation mechanism. The term "fractional-inspired" is used to emphasize that the weighting scheme reflects the long-memory characteristics of fractional calculus, without explicitly constructing a formal fractional differential operator. This explicit formulation ensures full reproducibility and removes ambiguity in the implementation of the proposed method.

Although memory-based mechanisms may resemble momentum-type ideas at a superficial level, fractional-inspired weighting differs in its capacity to flexibly capture longer-range dependencies in error evolution [19, 21]. By embedding this memory-aware term into the loss formulation, the framework links fractional intuition to measurable convergence behaviour in deep denoising. The experiments in Section 7 show that this loss-level composition reduces fluctuation intensity in the late training phase and yields smoother optimization trajectories under severe noise.

4. Proposed Application Framework

The proposed application framework is designed to isolate and evaluate the structural influence of memory-weighted (fractional-inspired) regularization on optimization dynamics in deep image denoising. Rather than embedding fractional concepts into network architecture or modifying optimization algorithms, the framework deliberately restricts intervention to the loss formulation. This design ensures that the computational graph, parameter update rules, and architectural components remain invariant across experiments, thereby eliminating confounding structural effects. A convolutional neural network (CNN)-based denoising model is adopted as a representative baseline, reflecting its widespread use in image restoration tasks. Network topology, parameter count, optimization strategy, learning rate schedule, and training duration are kept fixed throughout all experiments. Under this strictly controlled configuration, any variation in convergence smoothness, fluctuation intensity, or stability behaviour can be directly attributed to the modification of the loss structure and its associated weighting parameter λ. In this sense, the contribution of the framework lies not in architectural novelty but in methodological clarity. It establishes a transparent and controlled setting in which the impact of memory-aware loss construction on learning dynamics can be systematically examined.

A key component of the framework is the structural modification of residual aggregation prior to gradient computation. Memory-weighted regularization operates exclusively at the loss level, leaving forward propagation and backpropagation rules formally unchanged. The distinction emerges in how residual information is combined: isolated extreme deviations are attenuated, while persistent residual patterns receive proportionally greater emphasis. This reshaping of error aggregation alters the effective geometry of gradient formation and influences the smoothness and directionality of parameter updates, thereby modifying the optimization trajectory without introducing additional layers, parameters, or algorithmic complexity. The overall conceptual workflow of the proposed framework is illustrated in Figure 1, where the precise point of intervention is explicitly identified. This organization makes the point of intervention explicit and supports a single-variable study design in which only the loss construction (and λ) varies across runs.

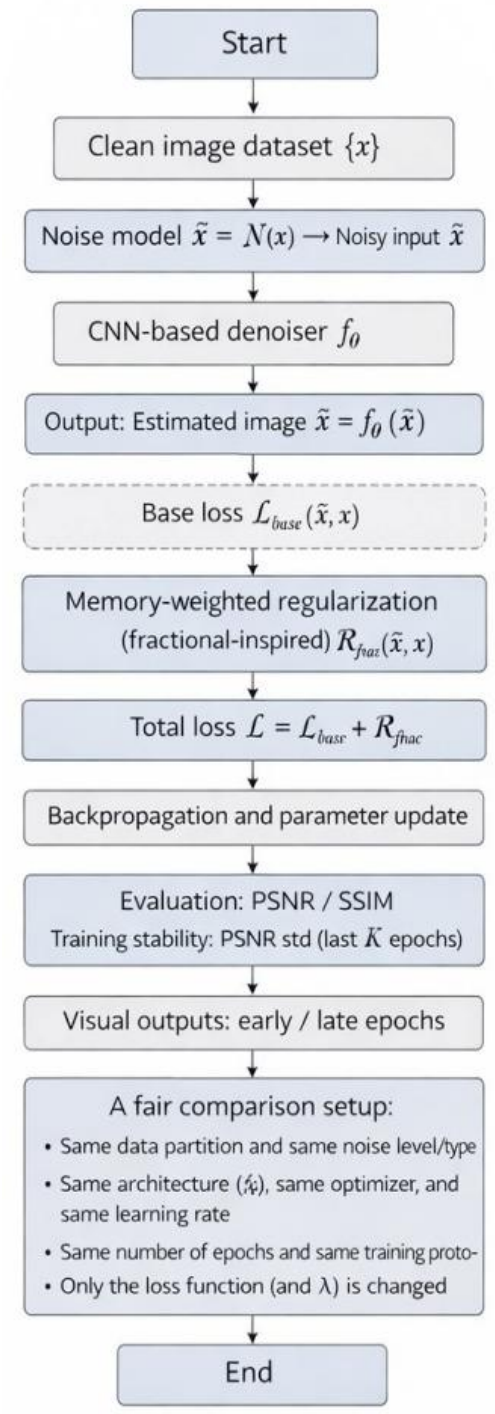

Fig. 1. Conceptual training workflow of the proposed framework, where the total loss is formed by combining the base loss, $\mathcal{L}_{base}$, with the memory-weighted (fractional-inspired) regularization term, $\mathcal{R}_{frac}$, while all architectural and training settings are kept fixed for fair comparison.

The training procedure adheres to the conventional supervised denoising setting in which noisy observations are derived from clean images, reconstructed outputs are produced by the network, and discrepancies are measured against ground truth references. The overall objective is formulated as a composite loss consisting of a base term $\mathcal{L}_{base}$ and a memory-weighted regularization component $\mathcal{R}$_frac, scaled by a coefficient λ (e.g., $\lambda \in [10^{-3}, 10^{-1}]$). While $\mathcal{L}_{base}$ enforces fidelity to instantaneous pixel-wise residuals, $\mathcal{R}$_frac modifies the aggregation of residual information across training iterations. As a result, the computational structure of gradient propagation remains unchanged, yet the effective optimization landscape is reshaped through the altered loss composition, influencing update smoothness and convergence behaviour. The training workflow and the precise loss-level intervention are illustrated in Figure 1.

The practical integration of memory-weighted regularization into the training loop is outlined in Table 1. Here, x̃ denotes the noisy input image, and x̂ represents the reconstructed output produced by the network.

This formulation characterizes the proposed mechanism as a structural modifier of gradient aggregation, operating at the level of loss construction while leaving the underlying optimization strategy intact. By preserving the computational graph and altering only the composition of the objective function, the framework establishes a controlled experimental setting in which the influence of memory-weighted regularization on convergence trajectories can be systematically evaluated. In particular, the modification reshapes the effective optimization landscape, moderates fluctuation intensity, and promotes enhanced stability across varying noise regimes.

Table 1. Memory-weighted regularization.

Input: Clean dataset $\mathcal{D} = \{x_i\}$, noise generator $\mathcal{N}(\cdot)$, where $\mathcal{N}(\cdot)$ denotes the noise corruption operator used to generate noisy observations from clean images, network model f_θ, base loss function $\mathcal{L}_{base}$, memory-weighted regularizer $\mathcal{R}_{frac}$, weight coefficient $\lambda > 0$, number of epochs E.
Output: Learned network parameters θ.

For e = 1 to E do:

1. Select a mini-batch $B \subset \mathcal{D}$.
2. Generate noisy inputs: $\tilde{x} \leftarrow \mathcal{N}(x), x \in B$.
3. Forward propagation: $\hat{x} \leftarrow f_\theta(\tilde{x})$.
4. Compute base loss: $\mathcal{L}_{base} \leftarrow \mathcal{L}_{base}(\hat{x}, x)$.
5. Compute memory-weighted term: $\mathcal{R} \leftarrow \mathcal{R}_{frac}(\hat{x}, x)$.
6. Total loss: $\mathcal{L} \leftarrow \mathcal{L}_{base} + \lambda\mathcal{R}$.
7. Backpropagation and parameter update: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$.

End for.

5. Problem Formulation and Noise Models

The supervised image denoising problem considered in this study assumes that a clean image x is not directly observable and that only a corrupted observation $\hat{x}$ is available [1, 2]. The degradation process is modelled as:

$$\hat{x} = x + n \tag{3}$$

where n denotes an additive noise component. A parameterized network model f_θ produces an estimate;

$$\hat{x} = f_\theta(\hat{x}) \tag{4}$$

and the model parameters are optimized by minimizing the discrepancy between $\hat{x}$ and the underlying clean image x. Within this formulation, the loss function serves as the central structural element governing how residual information is aggregated, how gradients are generated, and how the optimization trajectory evolves during training. Consequently, analysing denoising performance from a stability perspective requires examining the interaction between noise characteristics and loss-induced gradient behaviour.

Characterizing this interaction benefits from studying two complementary noise settings that stress loss behaviour in different ways. The first setting assumes additive Gaussian noise, modelled as $n \sim \mathcal{N}(0, \sigma^2)$, where increasing σ enlarges residual magnitudes in a broadly image-wide manner and yields errors that are comparatively diffused across pixels. When squared error-based objectives are used, this global scaling of residual energy can translate into proportionally larger gradient norms, which may increase sensitivity to noise intensity and promote oscillatory convergence as contamination grows [7, 17].

The second setting considers impulsive (salt-and-pepper) noise, in which a subset of pixels is driven to extreme intensity values with nonzero probability, producing sparse yet high-amplitude residuals and effectively heavy-tailed error behaviour. Since $L2$-type losses penalize large deviations quadratically, these isolated extremes can disproportionately shape the update direction and destabilize convergence trajectories [16, 17].

While robust losses attenuate the influence of extreme residuals, their effectiveness may vary when parameters remain fixed across different contamination levels. Together, these regimes establish a structured experimental setting in which magnitude scaling effects and outlier sensitivity can be examined concurrently. Within this dual-noise framework, the present study investigates how memory-weighted (fractional-inspired) regularization reshapes residual aggregation and influences convergence stability when integrated with both classical and robust loss formulations [6].

6. Experimental Setup

The experimental design is constructed to isolate and rigorously evaluate the structural influence of memory-weighted regularization on optimization dynamics within the image denoising task. All experiments follow a supervised training paradigm in which synthetically corrupted observations are generated from clean benchmark images and paired with their corresponding references. Standard datasets commonly used in the denoising literature are employed to ensure reproducibility and comparability. A fixed train–test partition is maintained across all experimental configurations, and identical noise-generation protocols are applied for each loss variant. This controlled setup ensures that any observed differences in reconstruction quality or convergence stability arise from the loss construction itself rather than from data sampling, distributional shifts, or architectural variations.

Throughout the study, architectural and optimization components remain strictly invariant. The same CNN-based denoising architecture, parameter count, optimization algorithm, learning rate schedule, batch size, and number of training epochs are employed for every loss configuration. Preserving the computational graph and training protocol across all settings eliminates confounding structural variability and ensures that any observed differences in convergence smoothness, fluctuation intensity, or reconstruction quality can be attributed directly to the structural properties of the loss formulation and its associated weighting parameter λ. Under this controlled configuration, the experimental framework functions as a diagnostic instrument for examining how modifications in residual aggregation alter gradient formation and reshape optimization trajectories without introducing additional architectural complexity.

Performance evaluation spans multiple intensity levels under both Gaussian and impulsive noise conditions, as defined in Section 5. Increasing noise intensity systematically elevates residual magnitudes and perturbs gradient responses, thereby establishing a structured stress-testing environment for analysing loss-induced convergence behaviour. Reconstruction accuracy is quantified using peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM), which remain standard and widely accepted metrics in image restoration research. However, the assessment extends beyond terminal performance values.

Convergence stability is explicitly incorporated as a primary evaluation dimension by computing the standard deviation of PSNR values over the final K training epochs. In all experiments, K is fixed to 5 epochs to quantify late-stage fluctuation behaviour under comparable conditions, and this fluctuation index captures the variability of optimization trajectories after initial transients have diminished. This combined reporting of reconstruction metrics and fluctuation-based stability indicators provides a comprehensive and structurally grounded view of both output quality and convergence dynamics across classical, robust, and memory-weighted loss formulations.

To improve reproducibility, all experiments are conducted under identical training conditions across different loss configurations. The same network architecture, optimizer, learning rate, batch size, and number of epochs are used for all methods. Furthermore, the results are evaluated over multiple noise realizations to ensure consistency of the observed behaviour. The reported metrics reflect stable trends rather than single-run observations

7. Results and Discussion

The comparative evaluation examines classical, robust, and memory-weighted loss structures with respect to reconstruction accuracy (PSNR, SSIM) and convergence stability. All results are obtained under the strictly controlled configuration described in Section 6, so differences can be attributed to the loss construction rather than architectural or optimizer variations. Alongside terminal PSNR/SSIM, the analysis emphasizes late-stage fluctuation behaviour to capture how each loss shapes the smoothness and consistency of the optimization trajectory.

Tables 2 and 3 summarize the quantitative findings. While the numerical differences in PSNR and SSIM appear relatively modest across the compared methods, it is important to interpret these results in the context of training dynamics rather than solely final reconstruction accuracy. In deep image denoising, small variations in PSNR often correspond to significant differences in convergence behaviour. In particular, the proposed memory-weighted formulation consistently reduces fluctuation during the late training phase, as reflected by the decrease in PSNR standard deviation. To further support this observation, we emphasize that stability is evaluated not only through final performance metrics, but also through fluctuation-based analysis computed over the final training window. The reduction in variance indicates improved convergence consistency and robustness against noise-induced oscillations. Although the absolute performance gains are moderate, the proposed method demonstrates a consistent trend across different noise conditions, indicating that the improvement is systematic rather than incidental. This highlights that the primary contribution of the proposed approach lies in enhancing training stability rather than maximizing peak reconstruction metrics. Incorporating memory-weighted regularization yields consistent improvements in PSNR and SSIM relative to both the classical $L2$ loss and the robust Cauchy formulation. Although the absolute numerical gains are moderate, their consistency indicates that the structural modification does not compromise reconstruction fidelity. More notably, the stability analysis reveals a clear reduction in convergence variability. The standard deviation of PSNR over the final K training epochs decreases from 0.011 under the $L2$ loss to 0.007 when memory-weighted regularization is employed. This reduction in late-stage fluctuation suggests that gradient updates evolve in a more coherent and less noise-sensitive manner, reflecting smoother optimization trajectories.

The primary objective of this study is not to achieve large improvements in reconstruction metrics, but to investigate how loss design influences convergence stability under noisy conditions. The numerical differences in PSNR and SSIM observed in Table 2 are relatively small across the compared methods. This behaviour is expected, as all models are trained under identical architectures and optimization settings, and differ only in the loss formulation. Therefore, the comparison should not be interpreted solely in terms of absolute reconstruction performance. A more informative perspective is provided by the fluctuation-based stability analysis presented in Table 3.

Table 2. PSNR/SSIM performance comparison.

Loss Structure	PSNR (dB)	SSIM
Squared error ($L2$)	26.59	0.9680
Robust loss (Cauchy)	26.56	0.9684
Robust loss with fractional regularization	**26.66**	**0.9687**

Table 3. PSNR stability comparison (last K epochs).

Loss Structure	PSNR (mean)	PSNR (std)
Squared error ($L2$)	27.29	0.011
Robust loss with fractional regularization	**27.42**	**0.007**

In particular, the reduction in PSNR standard deviation indicates improved consistency of the optimization trajectory. Although the absolute differences in standard deviation values may appear numerically small, they correspond to a meaningful reduction in late-stage oscillatory behaviour. In the context of deep learning optimization, such reductions in fluctuation are significant, as they reflect more stable gradient dynamics and improved robustness to noise-induced variations during training. This leads to more consistent reconstruction behaviour across iterations, even when the final PSNR values remain similar. Therefore, the results in Tables 2 and 3 should be interpreted from a stability-oriented perspective. The proposed memory-weighted formulation does not aim to produce large improvements in reconstruction metrics, but rather to enhance the consistency and smoothness of convergence. In relative terms, this corresponds to a reduction of approximately $30-40\backslash\%$ in fluctuation magnitude, which further highlights the practical significance of the observed improvement in convergence stability. This improvement in stability is particularly relevant for practical deployment, where consistent convergence behaviour is critical for reproducibility and robustness across varying noise conditions.

Visual evidence further supports the quantitative observations; however, the differences between methods should be interpreted with care. Figure 2 presents representative denoising examples obtained under different loss formulations, including L2, Cauchy, LogCosh, and the proposed memory-weighted loss. Within each panel, the images are arranged from top to bottom as the clean reference, the noisy input, and the denoised output. As shown in the figure, the visual differences among the compared methods remain relatively subtle in terms of overall perceptual quality. Nevertheless, the proposed memory-weighted formulation produces reconstructions with comparable visual fidelity while exhibiting slightly more consistent structural preservation and fewer localized fluctuations than the baseline configurations.

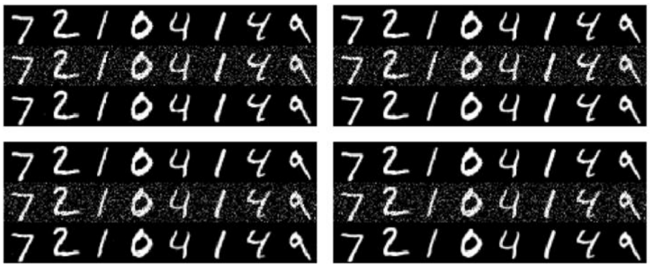

Fig. 2. Representative denoising examples obtained under different loss formulations.

In Figure 2, the four panels correspond to L2 loss (upper-left), Cauchy loss (upper-right), LogCosh loss (lower-left), and the proposed memory-weighted loss (lower-right). Within each panel, the images are arranged from top to bottom as the clean reference, the noisy input, and the denoised output. The visual differences among the methods are subtle, which is consistent with the fact that the primary contribution of the proposed approach lies in improving convergence stability rather than producing visually dominant changes. Accordingly, the figure is intended as a qualitative illustration, whereas the main evidence for improved stability is provided by the fluctuation-based quantitative analysis.

It is important to emphasize that the primary contribution of this work is not to achieve visually dominant improvements, but to enhance convergence stability through loss-level design. Therefore, the visual results should be interpreted as complementary qualitative evidence rather than standalone proof of superiority. When considered together with the quantitative analysis, particularly the reduction in PSNR standard deviation, these observations suggest that modifying residual aggregation improves the consistency of the reconstruction process rather than producing large perceptual differences.

The results demonstrate that memory-weighted loss composition yields measurable improvements in convergence stability while preserving reconstruction fidelity across challenging noise conditions. To further analyse the contribution of the memory-weighted term, an ablation perspective can be considered by comparing the base loss and the memory-augmented formulation under identical settings. The results indicate that incorporating the memory-weighted term consistently reduces fluctuation levels without degrading reconstruction quality, confirming its stabilizing effect.

8. Conclusion and Future Work

This study presents an application-oriented investigation of memory-weighted (fractional-inspired) regularization for stabilizing training dynamics in deep image denoising. The experimental results indicate that the proposed approach can improve convergence stability, particularly under high-noise conditions, by reducing fluctuation in the late training phase. While improvements in reconstruction metrics such as PSNR and SSIM are generally modest, the method demonstrates more consistent and smoother optimization behaviour compared to classical and robust loss functions.

Specifically, the study demonstrates that memory-weighted loss formulations can systematically reduce late-stage fluctuation in training dynamics under controlled experimental settings. However, several limitations should be noted. First, the observed performance gains remain relatively small in terms of absolute reconstruction metrics. Second, the experimental evaluation is conducted on a limited set of datasets and noise configurations. Third, the proposed formulation is not a strict fractional operator, but rather a fractional-inspired mechanism, which may limit its theoretical interpretation within the formal fractional calculus framework. Therefore, the findings of this study should be interpreted as preliminary but promising.

The results suggest that incorporating memory-aware structures into loss design can influence optimization dynamics, but further validation is required to fully establish its effectiveness across broader settings. This suggests that memory-aware loss design may serve as a complementary direction for improving optimization stability in deep learning models, particularly in noise-sensitive training scenarios within the scope of stability-oriented loss design.

Future work will focus on extending the proposed framework to larger-scale datasets, more complex architectures, and diverse noise models. In addition, a more rigorous theoretical analysis of the stability properties of memory-weighted loss functions will be investigated. Another important direction is the development of adaptive strategies for selecting the memory parameters, which may further enhance the practical impact of the approach.

Author Contributions

E.D. contributed to conceptualization, methodology, software, data curation, formal analysis, investigation, visualization, writing-original draft, and writing-review and editing. Y.Z. contributed to methodology, supervision, validation, and writing-review and editing. A.D. contributed to conceptualization, methodology, supervision, validation, and writing-review and editing. All authors have read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] B. Jiang, J. Li, Y. Lu, Q. Cai, H. Song, and G. Lu, "Efficient image denoising using deep learning: A brief survey," Inf. Fusion, vol. 118, p. 103013, 2025, doi: 10.1016/j.inffus.2025.103013.

[2] M. Elad, B. Kawar, and G. Vaksman, "Image denoising: The deep learning revolution and beyond—A survey

paper," SIAM J. Imaging Sci., vol. 16, no. 3, pp. 1594–1654, 2023, doi: 10.1137/23M1545859.

[3] A. Kaur and G. Dong, "A complete review on image denoising techniques for medical images," Neural Process. Lett., vol. 55, pp. 7807–7850, 2023, doi: 10.1007/s11063-023-11286-1.

[4] M. C. Sheeba and C. Seldev Christopher, "Adaptive deep residual network for image denoising across multiple noise levels in medical, nature, and satellite images," Ain Shams Eng. J., vol. 16, no. 1, p. 103188, 2025, doi: 10.1016/j.asej.2024.103188.

[5] A. Laghrib and A. Nachaoui, "A new fractional-order regularization for speckle image denoising: Preserving edges and features," Circuits Syst. Signal Process., vol. 44, pp. 3570–3598, 2025, doi: 10.1007/s00034-024-02981-y.

[6] R. R. Kumar and R. Priyadarshi, "Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches," Multimed. Tools Appl., vol. 84, pp. 10817–10875, 2025, doi: 10.1007/s11042-024-19313-6.

[7] F. Ullah, K. Kumar, T. Rahim, J. Khan, and Y. Jung, "A new hybrid image denoising algorithm using adaptive and modified decision-based filters for enhanced image quality," Sci. Rep., vol. 15, p. 8971, 2025, doi: 10.1038/s41598-025-92283-3.

[8] C. Yu, F. Ren, S. Bao, Y. Yang, and X. Xu, "Self-supervised ultrasound image denoising based on weighted joint loss," Digit. Signal Process., vol. 162, p. 105151, 2025, doi: 10.1016/j.dsp.2025.105151.

[9] M. Zubair, H. M. Rais, and T. Alazemi, "A novel attention-guided enhanced U-Net with hybrid edge-preserving structural loss for low-dose CT image denoising," IEEE Access, vol. 13, pp. 6909–6923, 2025, doi: 10.1109/ACCESS.2025.3526619.

[10] S. Paul, B. Jhamb, D. Mishra, and M. S. Kumar, "Edge loss functions for deep-learning depth-map," Mach. Learn. Appl., vol. 7, p. 100218, 2022, doi: 10.1016/j.mlwa.2021.100218.

[11] M. C. Kurucu, M. Güzelkaya, I. Eksin, and T. Kumbasar, "When fractional calculus meets robust learning: Adaptive robust loss functions," Knowl.-Based Syst., vol. 312, p. 113136, 2025, doi: 10.1016/j.knosys.2025.113136.

[12] O. Elharrouss, Y. Mahmood, Y. Bechqito, M. A. Serhani, E. Badidi, J. Riffi, and H. Tairi, "Task-based loss functions in computer vision: A comprehensive review," arXiv preprint arXiv:2504.04242, 2025.

[13] Q. Dao, K. Doan, D. Liu, T. Le, and D. Metaxas, "Improved training technique for latent consistency models," arXiv preprint arXiv:2502.01441, 2025.

[14] J. Gao, J. Shi, and F. Zhu, "Robust sparse unmixing via continuous mixed norm to address mixed noise," IEEE Geosci. Remote Sens. Lett., vol. 22, pp. 1–5, 2025, doi: 10.1109/LGRS.2025.3548697.

[15] J. Terven, D. M. Cordova-Esparza, J. A. Romero-González, A. A. López-Leyva, and J. A. Rodríguez-Reséndiz, "A comprehensive survey of loss functions and metrics in deep learning," Artif. Intell. Rev., vol. 58, p. 195, 2025, doi: 10.1007/s10462-025-11198-7.

[16] C. Chen, T. Liao, X. Deng, Z. Wu, S. Huang, and Z. Zheng, "Advances in robust federated learning: A survey with heterogeneity considerations," IEEE Trans. Big Data, vol. 11, no. 3, pp. 1548–1567, 2025, doi: 10.1109/TBDATA.2025.3527202.

[17] E. Ata, İ. O. Kıymaz, and H. M. Başkonuş, "Advanced fractional calculus approach to RC electrical circuit modeling: Analytical solutions and comparative behavioral analysis," Analog Integr. Circuits Signal Process., vol. 125, p. 32, 2025, doi: 10.1007/s10470-025-02511-z.

[18] E. Vieira-Martin, J. F. Gómez-Aguilar, J. E. Solís-Pérez, C. Escobar-Jiménez, and H. M. Baskonus, "Artificial neural networks: A practical review of applications involving fractional calculus," Eur. Phys. J. Spec. Top., vol. 231, no. 11, pp. 2059–2095, 2022, doi: 10.1140/epjs/s11734-022-00455-3.

[19] M. Joshi, S. Bhosale, and V. A. Vyawahare, "A survey of fractional calculus applications in artificial neural networks," Artif. Intell. Rev., vol. 56, pp. 13897–13950, 2023, doi: 10.1007/s10462-023-10474-8.

[20] O. Krejcar and H. Namazi, "Review of the applications of different fractional models in engineering," Appl. Math. Model., vol. 153, p. 116623, 2026, doi: 10.1016/j.apm.2025.116623.

[21] A. Lopes Ferrari, M. C. S. Gomes, A. C. R. Aranha, S. M. Paschoal, G. S. Matias, L. M. M. Jorge, and R. O. Defendi, "Mathematical modeling by fractional calculus applied to separation processes," Sep. Purif. Technol., vol. 337, p. 126310, 2024, doi: 10.1016/j.seppur.2024.126310.

[22] X. Zhang, D. Boutat, and D. Liu, "Applications of fractional operator in image processing and stability of control systems," Fractal Fract., vol. 7, no. 5, p. 359, 2023, doi: 10.3390/fractalfract7050359.

[23] S. Balochian and H. Baloochian, "Edge detection on noisy images using Prewitt operator and fractional order differentiation," Multimed. Tools Appl., vol. 81, pp. 9759–9770, 2022, doi: 10.1007/s11042-022-12011-1.

[24] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: A review," Artif. Intell. Rev., vol. 57, no. 11, 2024, doi: 10.1007/s10462-023-10631-z.

[25] O. Wu and R. Yao, "Data optimization in deep learning: A survey," IEEE Trans. Knowl. Data Eng., vol. 37, no. 5, pp. 2356–2375, 2025, doi: 10.1109/TKDE.2025.3530916.