Fast Trainable Primal-Dual TV Networks via Learnable Momentum Acceleration

Guangyu Yang
School of Automation, Qingdao University, Qingdao 266071, China

* Corresponding Author: **Guangyu Yang**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Impact Factor (RSIF): 8.40
Volume: 07
Issue: 01
Received: 16-11-2025
Accepted: 14-12-2025
Published: 12-01-2026
Page No: 95-105

Abstract
Deep unfolding networks based on total variation (TV) regularization provide an interpretable framework for image restoration by integrating model-driven optimization with data-driven learning. However, existing primal-dual unfolding networks often require a large number of iterative stages to achieve satisfactory restoration quality, leading to increased computational cost and slow convergence. To address this limitation, we propose a fast trainable primal-dual TV unfolding network with a learnable momentum acceleration mechanism. Building upon the classical primal-dual hybrid gradient optimization framework, an adaptive momentum strategy is introduced to accelerate information propagation across unfolding stages. Specifically, a dual-mechanism design combining exponential momentum decay and temporal fusion is developed to balance convergence speed and numerical stability. The resulting accelerated unfolding architecture retains the interpretability of variational optimization while significantly improving training and inference efficiency. Extensive experiments on image denoising benchmarks demonstrate that the proposed framework consistently achieves superior restoration performance with fewer unfolding stages and reduced runtime compared with existing TV-based and learning-based methods. The results verify the effectiveness of learnable momentum acceleration for efficient variational image restoration.

DOI: https://doi.org/10.54660/IJAIET.2026.7.1.95-105

Keywords: image restoration, total variation unfolding, primal-dual hybrid gradient, learnable momentum, accelerated convergence

1. Introduction

Image restoration aims to recover high-quality images from degraded observations and plays a fundamental role in numerous computer vision and computational imaging applications, including image denoising, image deblurring, compressed sensing reconstruction, and medical image analysis. Since image degradation is generally accompanied by information loss and noise contamination, image restoration is commonly formulated as an ill-posed inverse problem whose solution relies heavily on effective image priors and optimization strategies [1–3].

Variational methods have achieved remarkable success in image restoration owing to their solid mathematical foundations and interpretability. Among various regularization techniques, total variation (TV) regularization has become one of the most widely adopted image priors because of its capability to preserve edge structures while suppressing noise [4]. To further improve restoration performance, numerous extensions have been developed, including weighted TV models, nonlocal TV regularization, higher-order TV models, and total generalized variation (TGV) approaches [5–8]. In addition, the Fields-of-Experts (FoE) framework introduces trainable filter responses and nonlinear influence functions to characterize more complex image statistics, providing a flexible mechanism for learning image priors directly from data [9,10]. Although these methods have demonstrated favorable restoration performance, most

variational models still depend on iterative optimization procedures, which often require substantial computational effort for large-scale inverse problems.To efficiently solve variational restoration models, a variety of optimization algorithms have been

proposed, including proximal gradient methods, alternating direction method of multipliers (ADMM), operator splitting techniques, and primal-dual optimization schemes [11–14]. Among them, primal-dual hybrid gradient (PDHG) algorithms have attracted considerable attention because of their ability to efficiently handle nonsmooth optimization problems and their favorable convergence properties. More recently, variational neural networks and trainable primal-dual frameworks have demonstrated that iterative optimization algorithms can be transformed into learnable architectures, providing an effective bridge between model-based restoration and deep learning [15–18,21].

Deep unfolding has emerged as an important research direction for interpretable image restoration. By unfolding iterative optimization procedures into trainable neural network architectures, unfolding networks inherit the interpretability of variational models while simultaneously benefiting from the representation capability of deep learning. Representative examples include LISTA-based unfolding networks, ADMM-Net, ISTA-Net, trainable nonlinear reaction diffusion (TNRD) models, variational networks, and learned primal-dual frameworks [19–23]. Furthermore, recent image restoration methods based on convolutional neural networks, Transformer architectures, and foundation models have achieved remarkable restoration performance [24–28]. Nevertheless, these approaches often sacrifice interpretability or require substantial computational resources, making efficient and interpretable image restoration an active research topic.

Despite their promising restoration accuracy, the computational efficiency of unfolding networks remains a critical challenge. Existing primal-dual unfolding architectures generally employ a fixed number of unfolding stages and repeatedly update optimization variables throughout the network. As the number of stages increases, computational complexity, memory consumption, and inference latency grow accordingly. Moreover, the performance improvement obtained from additional unfolding stages gradually diminishes, resulting in an unfavorable efficiency-performance trade-off. To improve optimization efficiency, numerous acceleration strategies have been investigated in the optimization literature, including Nesterov acceleration [29], fast iterative shrinkage-thresholding algorithms (FISTA) [30], accelerated primal-dual schemes [31], momentum-based optimization methods [32], and adaptive restart acceleration techniques [33]. Although these methods have demonstrated remarkable success in improving convergence speed for large-scale optimization problems, most existing acceleration mechanisms are designed for conventional iterative algorithms and cannot be directly incorporated into trainable unfolding architectures.

Motivated by these observations, this paper proposes a Fast Primal-Dual Total Variation (FTPTV) network for efficient image restoration. Different from conventional primal-dual unfolding networks that rely solely on stage-wise iterative updates, the proposed method introduces a learnable momentum acceleration mechanism into the optimization process to enhance information propagation across unfolding stages. Specifically, adaptive momentum decay and temporal fusion strategies are jointly incorporated into the primal-dual optimization framework, resulting in accelerated convergence and improved computational efficiency. The resulting network preserves the interpretability of variational optimization while reducing redundant iterative computations. Extensive experiments on image denoising benchmarks demonstrate that the proposed framework achieves competitive restoration performance with significantly improved efficiency. The proposed framework is developed based on the trainable primal-dual variational network reported in our previous work [36].

The main contributions of this work are summarized as follows:

(1) The Fast Primal-Dual Total Variation (FTPTV) network is proposed for efficient image restoration, which transforms accelerated variational optimization into an interpretable trainable neural network architecture.

(2) A learnable momentum acceleration mechanism is introduced into the primal-dual optimization process to improve convergence speed and enhance information propagation across unfolding stages.

(3) Extensive experimental results on image denoising task demonstrate that the proposed framework achieves superior efficiency-performance trade-offs compared with existing variational and learning-based restoration methods.

2. Proposed Method

The proposed FTPTV network is developed based on the primal-dual hybrid gradient optimization framework. Different from conventional unfolding networks that rely solely on stage-wise iterative updates, the proposed method incorporates a learnable momentum acceleration mechanism into the unfolding process to improve convergence efficiency. Specifically, the variational restoration problem is first formulated within a total variation regularization framework. Subsequently, the corresponding primal-dual optimization procedure is unfolded into a trainable neural network architecture. To further enhance information propagation across unfolding stages, adaptive momentum decay and temporal fusion strategies are introduced into the optimization process. The overall framework preserves the interpretability of variational optimization while achieving improved computational efficiency. The main components of the proposed FTPTV network are described in the following subsections.

2.1. Variational Restoration Model

The proposed FTPTV network is derived from a variational image restoration model regularized by trainable total variation operators. Given a degraded observation f, the restoration process is formulated as an optimization problem consisting of a data-fidelity term and multiple trainable TV regularizers. Different from conventional TV models employing fixed differential operators, the proposed framework introduces learnable convolutional filters to characterize image structures at multiple scales and orientations.

The restoration energy is defined as

$$u = \underset{u}{\operatorname{argmin}} \frac{1}{2} \| Au - f \|_2^2 + \sum_{i=1}^{n} \gamma_i \| D_i u \|_1 \quad (1)$$

where u and f denote the latent image and degraded observation, respectively, A is the degradation operator, D_i represents the i-th trainable convolution operator, and γ_i denotes the corresponding regularization parameter. The collection of operators $(\|D_1u\|,\|D_2u\|,\cdots,\|D_nu\|)^T$ contains both first-order and higher-order anisotropic differential responses, allowing the model to simultaneously preserve edges, textures, and geometric structures.

The above variational formulation serves as the basis of the proposed FTPTV framework. To efficiently solve the optimization problem and simultaneously improve convergence efficiency, an accelerated primal-dual unfolding strategy is developed in the following subsection.

2.2. Accelerated Momentum Primal-Dual TV Unfolding

To efficiently solve the variational restoration model, we develop an accelerated momentum primal-dual unfolding framework. Different from conventional primal-dual unfolding networks, the proposed method introduces momentum propagation into the optimization process, enabling information reuse across adjacent unfolding stages and improving convergence efficiency.

At the k-th stage, an accelerated variable is first constructed using the current and previous image estimates

$$b^k = u^k + \rho_k\left(u^k - u^{k-1}\right) \tag{2}$$

where ρ_k denotes the momentum coefficient.

The dual variables are then updated as

$$p_i^{k+1} = p_i^k + D_i b^k \tag{3}$$

which incorporates the accelerated information into the dual optimization process.

Subsequently, the primal variable is updated by

$$u^{k+1} = u^k - A^T(Au^k - f) + \sum_{i=1}^{n} D_i^T D_i p_i^{k+1} \tag{4}$$

where D_i denotes the rotated operator corresponding to D_i.

To improve optimization stability, the momentum coefficient is adaptively learned during training according to

$$\rho_k = \alpha_k \rho_k^{base} \beta^k + (1-\alpha_k)\rho_{k-1} \tag{5}$$

where ρ_k^{base}, α_k, and β are trainable parameters.

The resulting optimization procedure consists of momentum propagation, dual updating, and primal updating, which are jointly unfolded into a trainable network architecture.

Compared with conventional primal-dual unfolding methods,
the proposed FTPTV framework effectively accelerates convergence while preserving the interpretability of variational optimization.

2.3. Design of Convolution and Nonlinear Functions

To improve the representation capability of the proposed FTPTV framework, trainable convolution operators and nonlinear functions are incorporated into the unfolding architecture.

For each operator D_i, the filter response is computed by $D_iu = k_i * u$ where k_i denotes the convolution kernel. To reduce the number of trainable parameters, the convolution kernel is represented as a linear combination of predefined basis filters

$$k_i = \sum_r \frac{\omega_{i,r} b_r}{\| \omega_i \|_2} \tag{6}$$

where b_r denotes the DCT basis and $\omega_{i,r}$ is the associated trainable coefficient.

This parameterization enables the network to learn directional and multi-scale image features while maintaining a compact model structure.

The regularization prior is constructed by aggregating nonlinear responses of multiple filter outputs. To address this issue, a

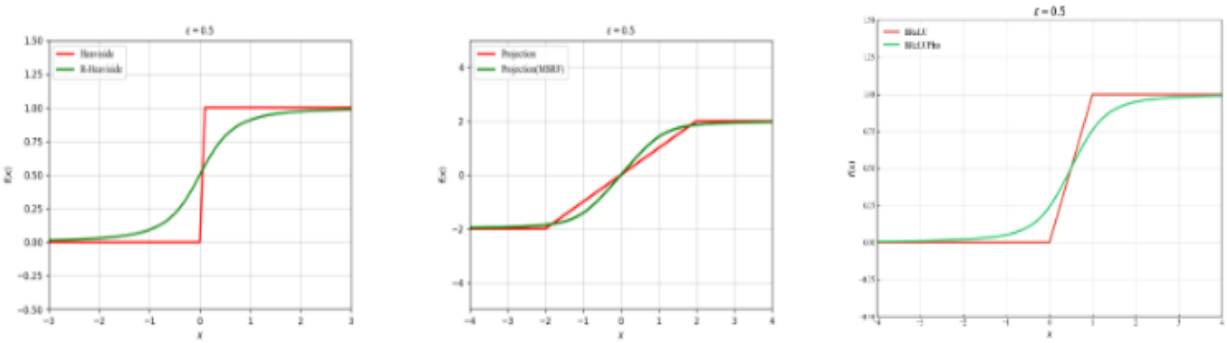
smooth approximation is introduced using a regularized Heaviside function which approximated by a sigmoid function,

$$\text{Projection}_{\lambda}(x) = \lambda H(x-\lambda) - \lambda(1-H(x+\lambda)) + x\big(H(x+\lambda) - H(x-\lambda)\big) \quad (7)$$

which maintains the projection behavior while providing a differentiable computational path.
Similarly, image normalization is implemented through the smooth BReLUPlus function

$$\text{BReLUPlus}(x) = \frac{1}{2}\left(1 + |x|_{\grave{o}} - |x-1|_{\grave{o}}\right) \quad (8)$$

where $|x|_{\grave{o}} = \sqrt{x^2 + \grave{o}}$, and $\grave{o}$ is a small positive parameter Compared with conventional piecewise activation functions, the proposed smooth formulation provides continuous derivatives and improves gradient propagation during training. The smooth approximations adopted in the proposed framework are illustrated in Fig. 1.
The convolution operators and nonlinear functions are jointly optimized together with the unfolding parameters, resulting in a compact and trainable FoE prior module for image restoration.



(a) The Heaviside function (b) The projection function (c) *BReLU function*

Fig 1: Smooth approximations of the Heaviside, projection and BReLU functions used in the proposed FTPTV framework.

2.4. Network Training

The proposed FTPTV network is trained in an end-to-end manner. All parameters involved in the unfolding framework are jointly optimized from training data, including the convolution kernels, nonlinear influence functions, primal-dual step sizes, and momentum coefficients.

Let u_T denote the output of the final unfolding stage and u_{gt} represent the corresponding ground-truth image. The network parameters are learned by minimizing the mean squared error (MSE) loss

$$\mathrm{L} = \frac{1}{N}\sum_{i=1}^{N}\left\|u_T^{(i)} - u_{gt}^{(i)}\right\|_2^2 \quad (9)$$

where N denotes the number of training samples.

The complete set of trainable parameters can be expressed as

$\Theta = \{k_i, \rho_i, \omega_i, \tau_k, \sigma_k, \rho_k\}$. Where k_i denotes the convolution kernels, ρ_i represents the nonlinear influence functions, ω_i denotes the kernel expansion coefficients, τ_k and σ_k are the primal-dual step sizes, and ρ_k is the momentum parameter introduced in the accelerated unfolding process.
During training, all parameters are optimized simultaneously through back-propagation. After convergence, the learned parameters are fixed and utilized for image restoration during inference. Owing to the proposed momentum propagation mechanism, the resulting FTPTV network achieves faster convergence while maintaining the interpretability of the underlying variational optimization model.
The architecture of one unfolding stage in the proposed FTPTV network is illustrated in Fig. 2. The detailed computational flow of momentum propagation and primal-dual optimization is shown in Fig. 3.

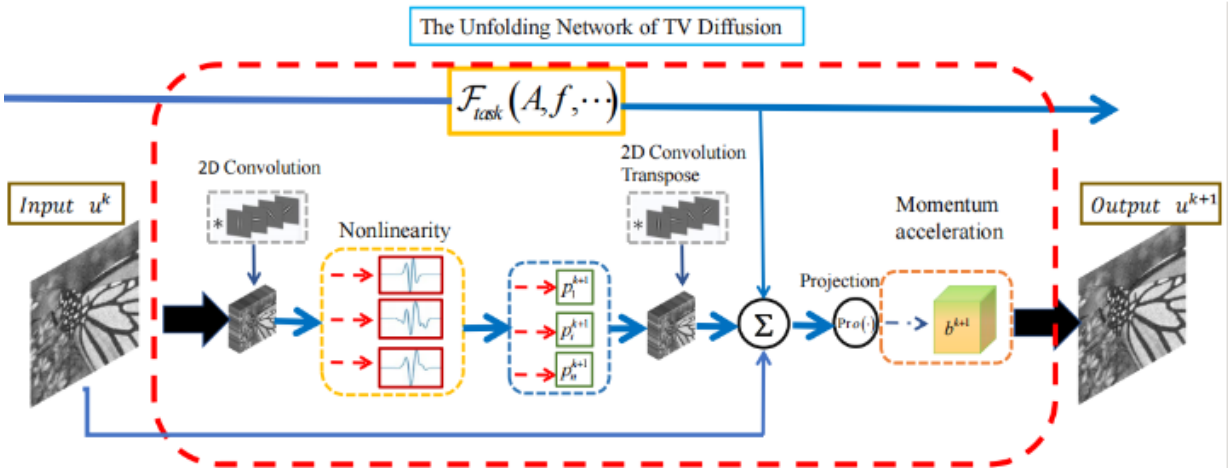


Fig. 2: Architecture of one unfolding stage in the proposed FTPTV network.

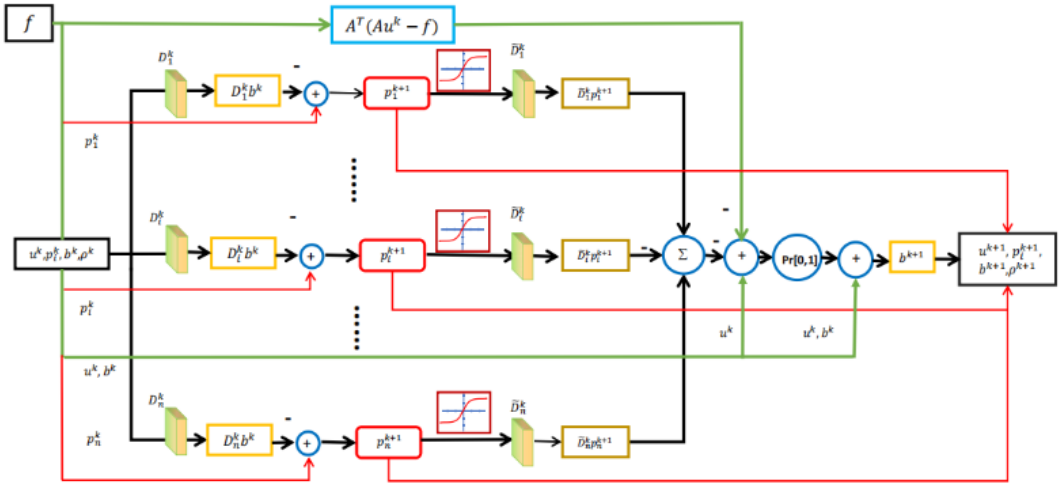


Fig 3: Detailed implementation of the accelerated primal-dual update process in FTPTV.

3. Experimental Results

3.1. Experimental Settings

To evaluate the effectiveness of the proposed Accelerated Momentum Primal-Dual Total Variation (FTPTV) network, experiments were conducted on both image denoising task. Quantitative and qualitative evaluations were performed to assess restoration accuracy, convergence behavior, and computational efficiency.

For image denoising, the BSD68 and CBSD68 benchmark datasets were adopted for grayscale and color image restoration, respectively. Following standard evaluation protocols, additive white Gaussian noise with noise levels of $\sigma = 15$, $\sigma = 25$, and $\sigma = 50$ was added to the clean images. Restoration quality was evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

The proposed FTPTV network was implemented using PyTorch and trained in an end-to-end manner. The number of unfolding stages was fixed to T=15 unless otherwise specified. The network parameters, including trainable convolution operators, nonlinear influence functions, primal-dual step sizes, and momentum coefficients, were jointly optimized using the Adam optimizer. The initial learning rate

was set to 10^{-4} and gradually decreased during training.

To investigate the contribution of the proposed acceleration mechanism, ablation experiments were first conducted on the momentum propagation strategy and smooth projection module. Subsequently, the proposed method was compared with representative variational and learning-based restoration approaches, including TV, TNRD, DnCNN, and the baseline primal-dual unfolding framework without momentum acceleration (TPTV).

In addition to restoration quality, convergence behavior and computational efficiency were also analyzed. The objective is to verify whether the introduced momentum acceleration mechanism can improve optimization efficiency while preserving the reconstruction capability and interpretability of the original primal-dual variational framework.

3.2. Ablation Study

To evaluate the effectiveness of the key components in the proposed FTPTV framework, ablation experiments were conducted on the momentum acceleration mechanism and the smooth projection strategy. The corresponding quantitative and visual results demonstrate that both modules contribute to the restoration performance and optimization stability of the proposed network.

3.2.1. Effect of Momentum Acceleration

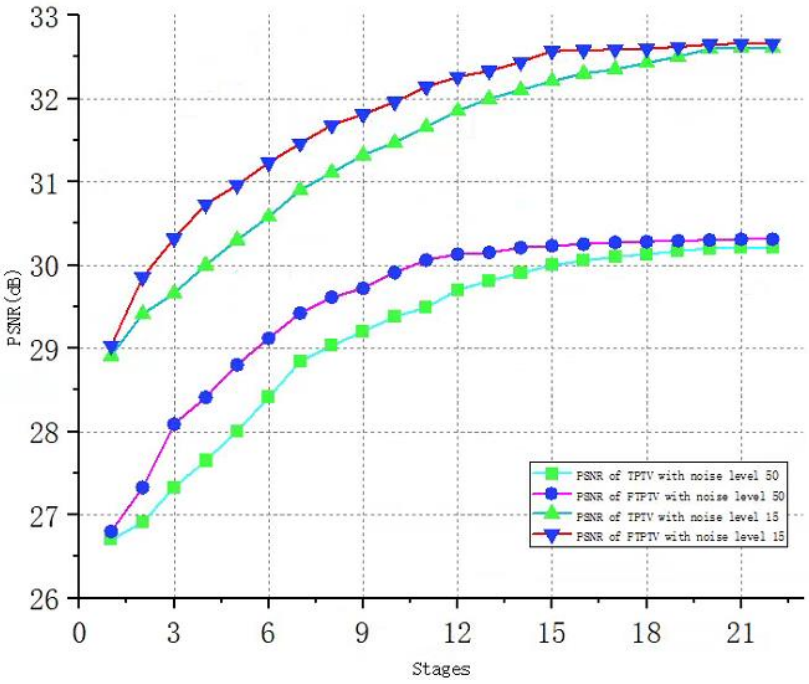


Fig 4: Performance comparisons of different stages in TPTV and FTPTV framework.

The effectiveness of the proposed momentum acceleration mechanism was evaluated by comparing the restoration performance and convergence behavior of the baseline primal-dual unfolding network and the accelerated FTPTV framework. The influence of the unfolding stages on restoration performance is illustrated in Fig. **4**.

As the number of unfolding stages increases, both models exhibit continuous performance improvements and gradually approach convergence. However, the accelerated framework consistently achieves higher PSNR values during the intermediate optimization stages. Benefiting from the momentum propagation mechanism, FTPTV reaches stable restoration performance with fewer unfolding stages, while the baseline primal-dual model requires additional iterations to obtain comparable results. The performance gain becomes marginal after approximately 15 stages, indicating that the proposed accelerated framework can achieve an effective trade-off between restoration accuracy and computational cost. Consequently, the unfolding depth is fixed to 15 stages in all subsequent experiments.

To further analyze the acceleration capability of the proposed momentum mechanism, the convergence epochs required to reach different target PSNR levels are presented in Fig. **5**. The proposed learnable momentum strategy consistently requires fewer training epochs than both the fixed-momentum and exponential-decay alternatives. This observation demonstrates that the adaptive momentum design effectively accelerates optimization and improves training efficiency. In particular, the performance advantage becomes increasingly evident at higher restoration quality levels, where conventional momentum strategies typically suffer from slower convergence.

3.2.2. Analysis of Learnable Momentum Parameter

To investigate the behavior of the learnable momentum parameter, the evolution of the momentum coefficient and its relationship with the iterative variables are illustrated in Fig. **6**.

As training progresses, the momentum parameter gradually decreases, reflecting the transition from aggressive exploration during the early stages to stable refinement in the later stages. Meanwhile, a consistent correlation can be observed between the momentum coefficient and the update ratio of the auxiliary and primal variables. Larger momentum values encourage faster information propagation and accelerate convergence during the initial optimization process, whereas smaller values help suppress oscillations near convergence and improve numerical stability.

These results indicate that the proposed learnable momentum mechanism is able to automatically adjust the optimization dynamics according to the current training state. By integrating adaptive momentum propagation into the primal-dual unfolding framework, the proposed FTPTV network achieves both accelerated convergence and stable restoration performance.

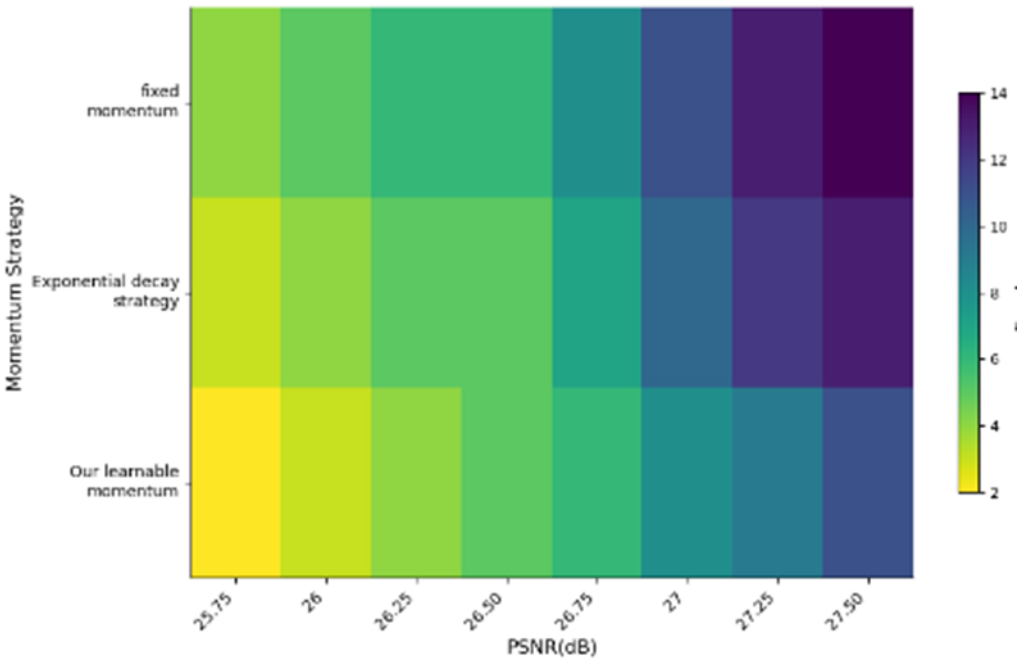


Fig 5: Convergence epochs for target PSNR at noise level 50

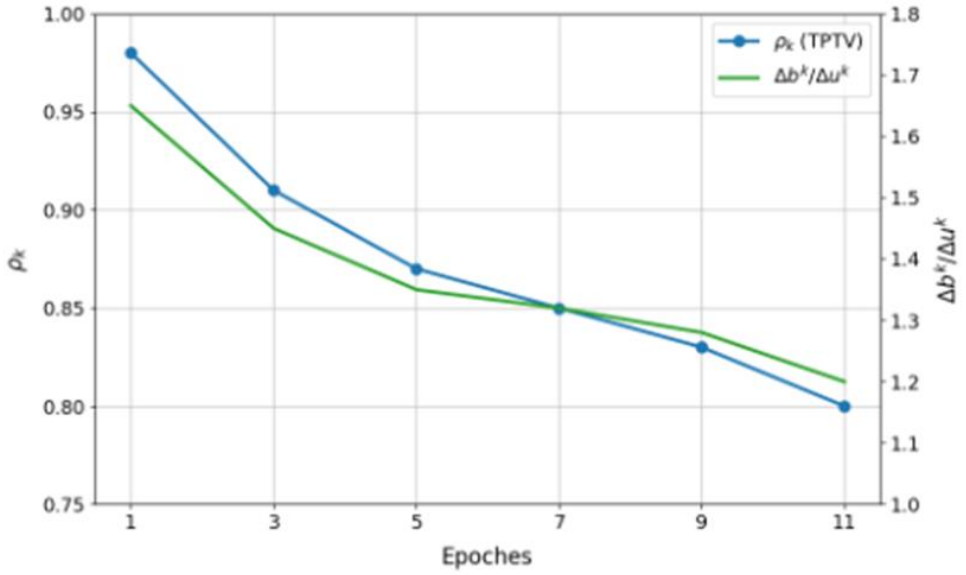


Fig 6: The adaptive correlation between the and coreiterative variables in the FTPTV network

3.3. Image Denoising Results

The restoration capability of the proposed FTPTV framework was evaluated on image denoising. Representative variational and learning-based restoration approaches were included for comparison. Quantitative evaluations were performed using PSNR and SSIM, while visual comparisons were conducted to assess the preservation of image structures and fine details.

To evaluate the denoising capability of the proposed unfolding networks, experiments were conducted on the BSD68 and CBSD68 benchmark datasets under different Gaussian noise levels. Quantitative results are summarized in Tables 1 and 2, while representative visual comparisons are presented in Figs. 7 and 8.

Table 1 reports the average PSNR and SSIM results on the BSD68 dataset. The compared methods include BM3D [34], TNRD [20], and the proposed unfolding frameworks. Both proposed methods consistently outperform the classical BM3D and TNRD approaches across all noise levels. Compared with TPTV, the momentum-accelerated FTPTV framework further improves restoration performance and achieves the highest PSNR values at all evaluated noise levels. In particular, FTPTV attains PSNR values of 31.74 dB, 29.21 dB, and 26.23 dB for noise levels $\sigma = 15$, 25, and 50, respectively. These results indicate that the introduced momentum acceleration mechanism improves optimization efficiency while preserving the restoration capability of the primal-dual unfolding framework.

The quantitative results on the CBSD68 dataset are shown in Table 2. Similar trends can be observed for color image denoising. Both TPTV and FTPTV achieve superior restoration performance compared with existing variational and learning-based

approaches. Although the performance differences between the two unfolding models are relatively small at low noise levels, FTPTV achieves comparable or superior performance across different noise levels and obtains the best result under severe noise corruption (σ=50).

Representative visual comparisons are shown in Fig. 7 for the Bridge image from BSD68. Compared with competing methods, the proposed unfolding networks preserve more structural details while effectively suppressing noise. In particular, the local texture patterns highlighted in the enlarged regions are reconstructed more faithfully by TPTV and FTPTV, whereas noticeable smoothing artifacts remain in BM3D and TNRD. The FTPTV model produces sharper edge structures and more natural visual appearance.

Fig. 8 presents color image denoising results on the Penguin image from CBSD68. Similar observations can be made in the highlighted regions. The proposed unfolding frameworks better preserve object boundaries and local texture information while reducing noise contamination. Benefiting from the accelerated optimization process, FTPTV achieves the best visual quality among the compared methods and generates clearer structural details around the penguin body and surrounding background regions.

Overall, both quantitative and qualitative evaluations demonstrate that the proposed TPTV and FTPTV frameworks effectively combine variational optimization and learnable image priors. Furthermore, the momentum-accelerated FTPTV model consistently provides improved restoration performance while requiring fewer unfolding stages, validating the effectiveness of the proposed acceleration mechanism for image denoising tasks.

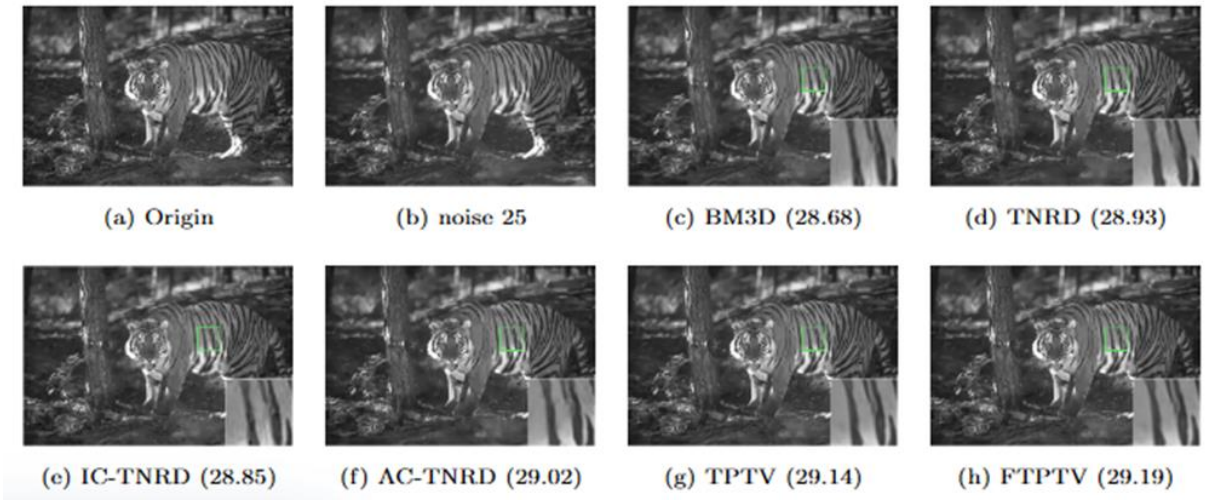


Fig 7: Gray image denoising images and PSNR(dB) on Bridge image from BSD 68 dataset for noise level 25.

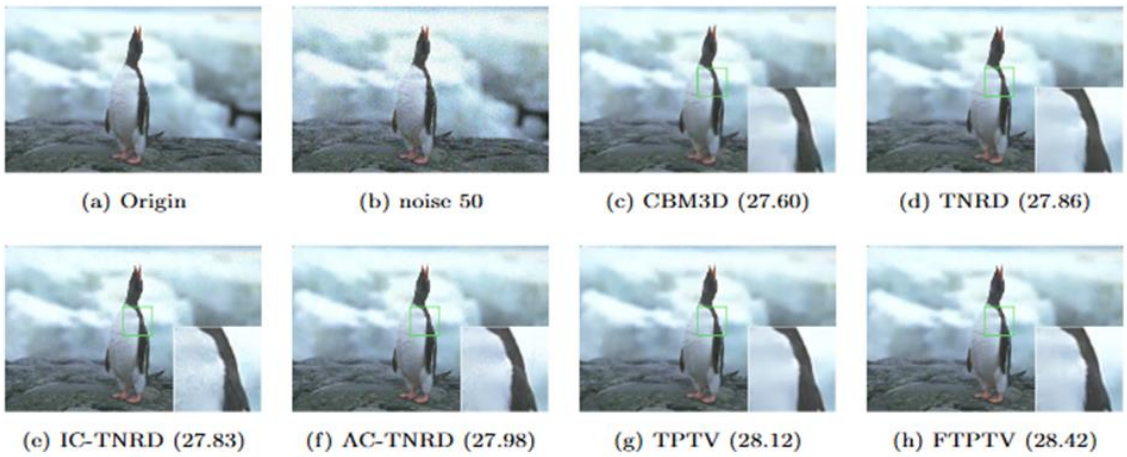


Fig 8: Color image denoising results and PSNR(dB) on the Penguin image from the CBSD68 dataset under noise level σ = 50.

Table 1: Average PSNR (dB) and SSIM results of image denoising on BSD68 dataset

Dataset	σ	BM3D (PSNR/SSIM)	TNRD (PSNR/SSIM)	TPTV (PSNR/SSIM)	FTPTV(PSNR/SSIM)
BSD68	15	31.08 / 0.872	31.42 / 0.883	31.71 / 0.885	31.74 / 0.887
BSD68	25	28.57 / 0.802	28.91 / 0.816	29.19 / 0.817	29.21 / 0.831
BSD68	50	25.62 / 0.687	25.96 / 0.702	26.21 / 0.725	26.23 / 0.725

Table 2: Average PSNR results of denoising on CBSD68 dataset.

Noise Level	CBM3D	TNRD	IRCNN	AC-TNRD	TPTV	FTPTV
σ = 15	33.31	33.79	33.87	33.84	33.90	33.92
σ = 25	30.53	31.07	31.18	31.11	31.22	31.19
σ = 50	27.42	27.80	27.88	27.82	28.05	28.20

3.4. Convergence and Efficiency Analysis

In addition to restoration accuracy, computational efficiency is an important consideration for practical image restoration applications. To evaluate the efficiency of the proposed FTPTV framework, model complexity and runtime performance were compared with several representative restoration methods. The quantitative results are summarized in Tables 3–5.

Table 3 reports the complexity comparison among different unfolding and deep learning-based models, including TNRD [20], DnCNN [24], TPTV, and the proposed FTPTV framework. The proposed FTPTV framework requires only 17K trainable parameters per unfolding stage, which is the smallest among all compared methods. The total number of trainable parameters is reduced from 420K in TPTV to 255K in FTPTV, while maintaining comparable restoration performance. Although DnCNN-B achieves the highest PSNR value, it requires substantially more parameters than the proposed unfolding framework. These results indicate that FTPTV provides a more efficient parameter utilization strategy through the momentum-accelerated unfolding architecture.

The runtime comparison on the Set12 dataset with Gaussian noise level σ = 50 is presented in Table 4. Traditional optimization-based algorithms, including TV-PD and TV-PDHG, require a large number of iterative updates and therefore incur higher computational costs. By introducing momentum acceleration, FTPTV reduces the unfolding depth from 20 stages in TPTV to 15 stages while achieving slightly higher restoration performance. Consequently, the CPU runtime decreases from 6.78 s to 4.42 s, and the GPU runtime decreases from 0.151 s to 0.083 s. At the same time, FTPTV achieves the highest PSNR and SSIM values among the compared methods.

To further evaluate scalability, Table 5 presents the runtime performance under different image resolutions. As image size increases, the computational cost of all methods increases accordingly. Nevertheless, FTPTV consistently exhibits the shortest execution time among the learning-based restoration models. For images with a resolution of 1024 × 1024, FTPTV requires only 4.42 s on CPU and 0.083 s on GPU, compared with 6.78 s and 0.151 s for TPTV, respectively. Similar improvements can also be observed for lower-resolution images.

Table 3: Complexity comparison with different methods

Method (Layers)	TNRD (5)	IC-TNRD (5)	AC-TNRD (5)	DnCNN-B (20)	TPTV (20)	FTPTV (15)
Parameters per Layer	40K	29K	41K	40K	21K	17K
Total Parameters	200K	145K	202K	665K	420K	255K
PSNR (dB)	26.81	25.77	26.88	27.21	27.14	27.17

Table 4: Runtime comparison with conventional TV-PD/PDHG solvers on Set12 under Gaussian noise σ = 50.

Method	Iterations / Layers	PSNR (dB)	SSIM	CPU Time (s)	GPU Time (s)
TV-PD	100	26.31	0.752	18.52	–
TV-PDHG	80	26.55	0.768	12.85	–
TPTV	20	27.14	0.792	6.78	0.151
FTPTV	15	27.17	0.795	4.42	0.083

Table 5: Runtime comparison for image denoising with different models under different image resolutions. BM3D is evaluated on CPU, while the learning-based models are reported withCPU/GPU time.

Resolution	BM3D	DnCNN-S (17)	DnCNN-B (20)	TPTV (20)	FTPTV (15)
256 × 256	0.65	0.74 / 0.014	0.90 / 0.016	0.60 / 0.020	0.47 / 0.011
512 × 512	2.85	3.41 / 0.051	4.11 / 0.060	2.11 / 0.050	1.66 / 0.032
1024 × 1024	11.89	12.1 / 0.200	14.1 / 0.235	6.78 / 0.151	4.42 / 0.083

Overall, the results in Tables 3–5 demonstrate that the proposed FTPTV framework achieves an effective balance between restoration performance, model complexity, and computational efficiency. By incorporating momentum acceleration into the primal-dual unfolding architecture, FTPTV significantly reduces computational overhead while maintaining competitive restoration quality.

4. Conclusion

The experimental results demonstrate that the proposed FTPTV framework effectively combines the interpretability of variational optimization with the efficiency of momentum-accelerated unfolding. Compared with the baseline TPTV framework, the introduced momentum propagation mechanism significantly improves convergence behavior while maintaining competitive restoration performance across denoising task.

Another important observation is that the acceleration strategy introduces only a small number of additional trainable parameters, yet provides noticeable improvements in optimization efficiency. Meanwhile, the smooth projection mechanism ensures stable end-to-end training and preserves the constraint-enforcement capability inherited from the original primal-dual optimization framework. These results indicate that the proposed FTPTV framework provides an effective balance between restoration

accuracy, convergence speed, and computational efficiency for image restoration applications.
Future work will investigate adaptive unfolding depth design,
Mixture-of-Experts based variational regularization,
and the integration of large-scale vision foundation models
into trainable variational unfolding frameworks.

5. References

1. Bertero M, Boccacci P. Introduction to inverse problems in imaging. Bristol: Institute of Physics Publishing; 1998.
2. Chan TF, Shen J. Image processing and analysis: variational, PDE, wavelet, and stochastic methods. Philadelphia: SIAM; 2005.
3. Hansen PC. Discrete inverse problems: insight and algorithms. Philadelphia: SIAM; 2010.
4. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D. 1992;60(1-4):259-68.
5. Bredies K, Kunisch K, Pock T. Total generalized variation. SIAM J Imaging Sci. 2010;3(3):492-526.
6. Gilboa G, Osher S. Nonlocal operators with applications to image processing. Multiscale Model Simul. 2008;7(3):1005-28.
7. Dong W, Zhang L, Shi G, Li X. Nonlocally centralized sparse representation for image restoration. IEEE Trans Image Process. 2013;22(4):1620-30.
8. Lefkimmiatis S, Bourquard A, Unser M. Hessian-based norm regularization for image restoration with biomedical applications. IEEE Trans Image Process. 2012;21(3):983-95.
9. Roth S, Black MJ. Fields of experts: a framework for learning image priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2005. p. 860-7.
10. Schmidt U, Roth S. Shrinkage fields for effective image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2014. p. 2774-81.
11. Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting. Multiscale Model Simul. 2005;4(4):1168-200.
12. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn. 2011;3(1):1-122.
13. Condat L. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J Optim Theory Appl. 2013;158(2):460-79.
14. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vis. 2011;40(1):120-45.
15. Monga V, Li Y, Eldar YC. Algorithm unrolling: interpretable, efficient deep learning for signal and image processing. IEEE Signal Process Mag. 2021;38(2):18-44.
16. Gregor K, LeCun Y. Learning fast approximations of sparse coding. In: Proceedings of the International Conference on Machine Learning (ICML); 2010. p. 399-406.
17. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. In: Advances in Neural Information Processing Systems (NeurIPS); 2016.
18. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, *et al*. Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med. 2018;79(6):3055-71.
19. Zhang J, Ghanem B. ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018. p. 1828-37.
20. Chen Y, Pock T. Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration. IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1256-72.
21. Adler J, Öktem O. Learned primal-dual reconstruction. IEEE Trans Med Imaging. 2018;37(6):1322-32.
22. Lefkimmiatis S. Universal denoising networks: a novel CNN architecture for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018. p. 3204-13.
23. Zhang K, Zuo W, Gu S, Zhang L. Learning deep CNN denoiser prior for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 3929-38.
24. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans Image Process. 2017;26(7):3142-55.
25. Zhang K, Li Y, Zuo W, Zhang L, Van Gool L, Timofte R. Plug-and-play image restoration with deep denoiser prior. IEEE Trans Pattern Anal Mach Intell. 2022;44(10):6360-76.
26. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, *et al*. Restormer: efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022. p. 5728-39.
27. Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. In: Proceedings of the European Conference on Computer Vision (ECCV); 2022. p. 17-33.
28. Potlapalli V, Zamir SW, Khan S, Khan FS. PromptIR: prompting for all-in-one image restoration. In: Advances in Neural Information Processing Systems (NeurIPS); 2023.
29. Nesterov Y. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Sov Math Dokl. 1983;27:372-6.
30. Beck A, Teboulle M. A fast-iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2009;2(1):183-202.
31. Chambolle A, Pock T. On the ergodic convergence rates of a first-order primal-dual algorithm. Math Program. 2016;159:253-87.

32. Polyak BT. Some methods of speeding up the convergence of iteration methods. USSR Comput Math Math Phys. 1964;4(5):1-17.
33. O'Donoghue B, Candès E. Adaptive restart for accelerated gradient schemes. Found Comput Math. 2015;15(3):715-32.
34. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process. 2007;16(8):2080-95.
35. Dabov K, Foi A, Katkovnik V, Egiazarian K. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. In: Proceedings of the IEEE International Conference on Image Processing (ICIP); 2007. p. 313-6.
36. Yang G, Wei W, Pan Z. A variational neural network for image restoration based on coupled regularizers. Multimed Tools Appl. 2024;83:12379-401.

How to Cite This Article
Yang G. Fast trainable primal-dual TV networks via learnable momentum acceleration. Int J Artif Intell Eng Transform. 2026;7(1):95–105. doi:10.54660/IJAIET.2026.7.1.95-105.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.