

Physics-Informed Image Restoration via Progressive PDE Integration

Shamika Likhite*, Santiago López-Tapia, Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering

Northwestern University, Evanston, IL, USA

shamikalikhite2024@u.northwestern.edu, {santiago.lopeztapia, a-katsaggelos}@northwestern.edu

Abstract—Motion blur, caused by relative movement between camera and scene during exposure, significantly degrades image quality and impairs downstream computer vision tasks such as object detection, tracking, and recognition in dynamic environments. While deep learning-based motion deblurring methods have achieved remarkable progress, existing approaches face fundamental challenges in capturing the long-range spatial dependencies inherent in motion blur patterns. Traditional convolutional methods rely on limited receptive fields and require extremely deep networks to model global spatial relationships, while transformer-based approaches with self-attention mechanisms can model global context but introduce additional architectural complexity. These limitations motivate the need for alternative approaches that incorporate physical priors to guide feature evolution during restoration. In this paper, we propose a novel motion deblurring framework that integrates physics-informed partial differential equation (PDE) dynamics into state-of-the-art restoration architectures through a progressive training strategy. By leveraging advection-diffusion equations to model feature evolution, our approach naturally captures the directional flow characteristics of motion blur while enabling principled global spatial modeling. The advection component directs information flow along image structures to preserve edges and textures, while the diffusion component provides adaptive smoothing to reduce artifacts during the restoration process. Our PDE-enhanced deblurring models achieve superior restoration quality with minimal overhead, adding only approximately 1% to inference GMACs (0.6 to 2.4 GMACs depending on the base architecture) while providing consistent improvements in perceptual quality across multiple state-of-the-art architectures. Comprehensive experiments on standard motion deblurring datasets (GoPro, RealBlur-R, and RealBlur-J) demonstrate that our physics-informed approach improves PSNR and SSIM significantly across four diverse architectures, including FFTransformer, NAFNet, Restormer, and Stripformer. These results validate that incorporating mathematical physics principles through PDE-based global layers can enhance deep learning-based image restoration, establishing a promising direction for physics-informed neural network design in computer vision applications.

1 INTRODUCTION

Motion blur significantly degrades image quality and reduces accuracy in subsequent computer vision tasks such as object classification, scene understanding, and activity recognition. For instance, analyzing high-speed photography for motion capture or automated analysis is crucial in sports analytics and surveillance applications. Such dynamic imaging scenarios inherently involve rapid subject movements and camera operations, introducing substantial motion blur that severely compromises the accuracy of object detection, feature extraction, and recognition systems. Similarly, mobile photography and handheld camera

systems frequently encounter motion blur due to camera shake, subject movement, or inadequate stabilization, resulting in degraded image quality that significantly impacts downstream computer vision applications in real-world scenarios.

A fundamental challenge in image deblurring is effectively capturing long-range spatial dependencies in motion blur patterns. Traditional convolutional approaches rely on limited receptive fields and require very deep networks to model global spatial relationships, while recent transformer-based methods achieve global modeling through self-attention mechanisms but with additional architectural complexity. Current state-of-the-art image restoration methods typically employ data-driven optimization without explicitly incorporating physical priors about the blur formation and restoration process. This motivates exploring physics-informed approaches that can guide feature evolution through mathematical constraints while maintaining restoration quality.

Inspired by the work "Condensing CNNs with Partial Differential Equations," [19], we propose a novel methodology that incorporates PDE-informed principles into image restoration networks. Unlike traditional approaches that rely solely on data-driven optimization, our method leverages the mathematical foundations of differential equations to model feature evolution during the restoration process. Specifically, we introduce PDE-based global layers that capture the directional flow characteristics inherent in motion blur patterns through advection-diffusion dynamics. To address the gradient stability challenges that arise when integrating iterative PDE solvers into deep networks, we develop a progressive training strategy that gradually increases the number of PDE iterations while correspondingly reducing the temporal step size to maintain constant total integration time. This progressive approach enables stable convergence while allowing the network to benefit from finer temporal discretization as training progresses. By incorporating mathematical priors that model blur formation as a physical process, we achieve improved restoration quality across multiple state-of-the-art architectures with minimal additional computational overhead.

In this paper, our main contributions are:

- 1) We propose the incorporation of a Neural Network layer inspired by the advection-diffusion PDE into deep learning image restoration models through a progressive training strategy. This allows feature evolution to be guided by advection-diffusion dynamics, enabling more effective restoration while preserving global spatial interactions.
- 2) We show that incorporating our training method into

* denotes corresponding author

existing architectures leads to substantial improvements in motion deblurring, enhancing perceptual quality and downstream vision task performance without significantly increasing computational complexity.

2 RELATED WORK

A. Evolution of Image Deblurring Methods

Image deblurring has evolved significantly from early traditional approaches to sophisticated deep learning methods that leverage advanced architectural designs. Traditional deconvolution-based methods [1,2] rely on mathematical models of the blur process, often assuming uniform blur kernels and struggling with complex, spatially-varying motion patterns. These approaches suffer from computational inefficiency and suboptimal restoration quality due to their reliance on hand-crafted priors and simplified blur models. Modern image deblurring methods can be broadly categorized into single-image and multi-scale approaches. Single-image methods [3,4] process individual blurred images directly, focusing on learning effective representations for blur removal. Multi-scale methods [5,6] exploit hierarchical feature representations at different resolutions, enabling better handling of large blur kernels and complex motion patterns but often at the cost of increased computational overhead. The advent of deep learning has revolutionized image restoration tasks. Convolutional Neural Networks (CNNs) have been extensively employed for image deblurring, with architectures ranging from simple encoder-decoder networks to complex multi-scale designs [7,8]. Recent works have explored attention mechanisms [9] and transformer architectures [10] to better model long-range spatial dependencies crucial for effective blur removal. NAFNet [11] introduced a novel approach by simplifying network architectures while maintaining competitive performance, demonstrating that carefully designed simple blocks can be highly effective for image restoration tasks. This work highlighted the importance of architectural efficiency and inspired subsequent research into lightweight yet powerful restoration models, reinforcing the principle that architectural simplicity can coexist with high performance in image deblurring applications.

B. Advanced Architectures and Spatial Modeling

Effective spatial feature modeling remains a fundamental challenge in image restoration. Traditional CNN-based methods [13,14] rely on local convolutions with limited receptive fields, requiring deep networks to capture long-range spatial dependencies necessary for handling large blur kernels. Dilated convolutions [15,16] offer an alternative by expanding receptive fields without increasing computational cost, but they can introduce artifacts and struggle with irregular blur patterns. Vision transformers have gained significant attention for their ability to model long-range dependencies effectively [22,23]. In image processing, transformers have been applied to capture global spatial relationships, often outperforming CNN-based approaches on various restoration tasks [24,25]. However, the quadratic complexity of self-attention mechanisms poses challenges for high-resolution image processing. Recent works

have explored efficient transformer variants and hybrid CNN-transformer architectures to mitigate computational overhead while preserving the modeling capabilities of attention mechanisms [26,27].

C. Mathematical Foundations and Training Methodologies

The integration of Partial Differential Equations (PDEs) with neural networks has emerged as a promising research direction. The work "Condensing CNNs with Partial Differential Equations" [19] demonstrated how PDE principles can be incorporated into CNN architectures to achieve better parameter efficiency and improved performance in classification tasks. This approach leverages the mathematical foundations of differential equations to guide feature evolution, resulting in more principled network designs. PDE-informed methods have shown particular promise in capturing continuous spatial relationships, making them well-suited for image processing tasks where understanding spatial dependencies is crucial. However, their application to image restoration problems remains largely unexplored.

While various efficient architectures have been proposed for image restoration [20], [21], there is growing interest in approaches that enhance restoration quality through physics-informed priors rather than solely focusing on architectural efficiency. Methods that incorporate mathematical constraints can potentially improve performance by leveraging domain knowledge about the degradation and restoration processes. The challenge lies in effectively integrating such physics-based components into existing deep learning architectures while maintaining stable training dynamics and restoration quality.

D. Motivation and Challenges

Despite significant progress in image deblurring, current approaches face challenges in effectively modeling long-range spatial dependencies. Traditional CNN-based methods rely on limited receptive fields and require very deep networks to capture global spatial relationships, while transformer-based approaches achieve global modeling through self-attention mechanisms at the cost of additional architectural complexity. Most existing methods employ purely data-driven optimization without explicitly incorporating physical priors about blur formation and restoration processes.

While Kag et al. [19] demonstrated that PDE-based global layers can be effectively integrated into CNNs for classification tasks, their approach has not been explored for image restoration problems. Image restoration presents unique challenges compared to classification, including spatially varying degradations, the need to preserve fine-grained details and textures, perceptual quality requirements that go beyond pixel-level metrics, and the need for global spatial coherence during reconstruction. More importantly, we have determined experimentally that using PDE steps inside a restoration network brings new stability issues. We hypothesize that, since the solver runs many steps in a row, backpropagation must pass through a long chain, which can make gradient calculation unstable, similar to recurrent neural networks.

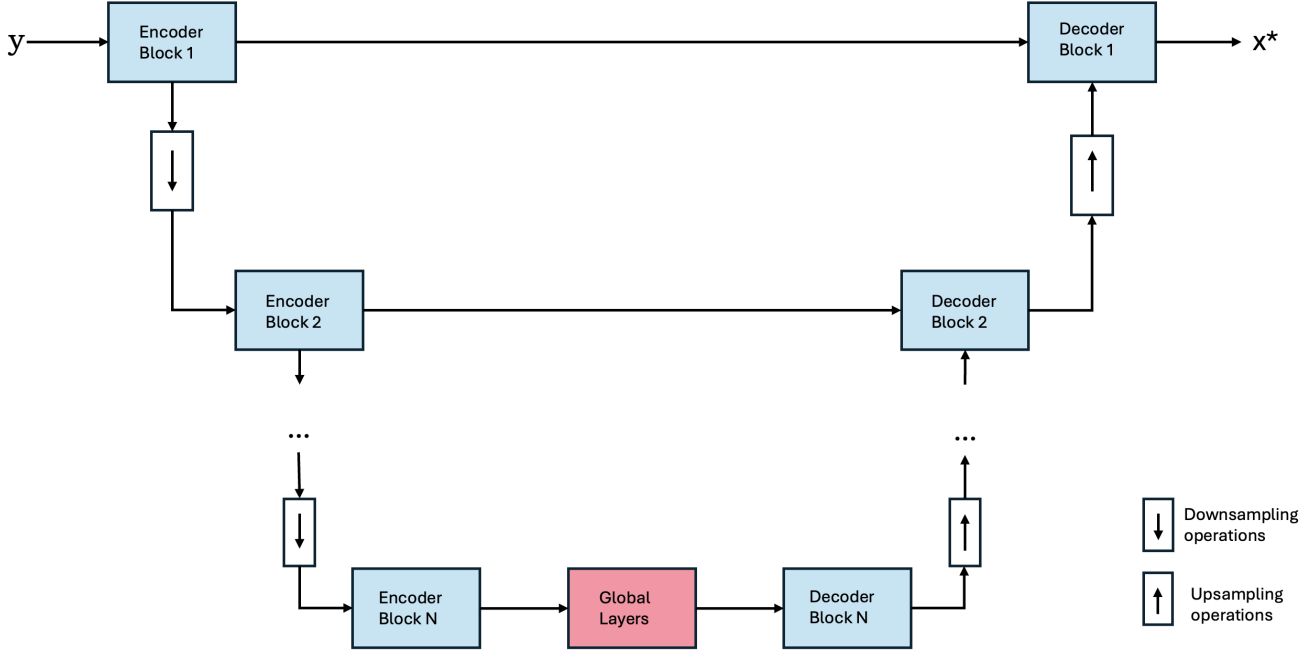


Fig. 1. High-level architecture of the image restoration network. The network follows a U-Net style encoder-decoder structure, where the input degraded image y is progressively downsampled through encoder blocks (Encoder Block 1 to N) via max pooling operations. At the bottleneck, global layers process the latent representation at the coarsest scale. The decoder path progressively upsamples the features through decoder blocks (Decoder Block N to 1) using upward convolutions to reconstruct the restored image x^* . Skip connections between corresponding encoder and decoder blocks preserve spatial information across scales. The global layers at the bottleneck serve as the primary location for integrating physics-informed processing.

3 METHOD

Our approach leverages Partial Differential Equations (PDEs) in a Global Feature Layer [19] that achieves efficient global spatial modeling for image deblurring. Instead of relying on deep networks or large kernels to capture long-range dependencies, we embed an advection-diffusion PDE solver that processes features through lightweight iterative operations. As demonstrated in [19], the use of these layers enables global receptive fields with linear complexity $O(K \times H \times W)$, with K being the number of iterations of the discretized PDE. This reduces computational overhead compared to traditional CNN approaches (where complexity increases with the size of the receptive field) while maintaining superior restoration quality through physics-informed feature processing.

A. Global Feature Layer and PDE Formulation for Image Restoration

The Global Feature Layer implements a physics-inspired approach to image restoration where the discretized PDE is used to construct the layer, constraining the output feature map H following the framework established in [19]. To guide the training of our restoration network, the underlying dynamics are modeled using the advection-diffusion equation:

$$\frac{\partial H}{\partial t} = \nabla \cdot (D \nabla H) - \nabla \cdot (\mathbf{v} H) + f(I), \quad (1)$$

where D represents the diffusion coefficient(s), \mathbf{v} denotes the velocity field, and $f(I)$ is a source term depending on the input features I .

This PDE interprets the restoration process as **guided particle dynamics** where input pixels evolve under physical constraints:

- **Diffusion term** ($\nabla \cdot (D \nabla H)$): Models smoothing where particles move from high-concentration to low-concentration regions. The diffusion coefficient D controls the strength of this spreading. This term handles noise reduction and fills gaps, but without guidance would blur all structures uniformly.
- **Advection term** ($\nabla \cdot (\mathbf{v} H)$): Directs information flow along image structures. The velocity field $\mathbf{v} = (u, v)$ determines the direction and speed of particle movement. This allows information to propagate along edges and textures rather than spreading equally in all directions.
- **Source term** ($f(I)$): Injects task-specific information from the degraded input I , anchoring the solution to observed data and guiding the restoration process.

Blur/Deblur Modeling

The advection and diffusion terms work together to enable structure-aware restoration. While the diffusion term smooths the feature map, the velocity field \mathbf{v} guides how this smoothing occurs by directing flow along image structures. This allows the model to smooth homogeneous regions while maintaining

sharpness along edges and textures. In deblurring, the velocity field guides pixel movement along structural boundaries, helping recover sharp details without introducing artifacts or over-smoothing.

1) 2D Image Domain

Since images are inherently 2D, we expand the velocity and diffusion terms into spatial components: $\mathbf{v} = (u, v)$ and $D = (D_x, D_y)$. Following the approach in [19], Eq.(1) expands to:

$$\begin{aligned} \frac{\partial}{\partial t} H(x, y, t) + \frac{\partial}{\partial x} (u(x, y, t) H(x, y, t)) + \frac{\partial}{\partial y} (v(x, y, t) H(x, y, t)) \\ = \frac{\partial}{\partial x} \left(D_x \frac{\partial}{\partial x} H(x, y, t) \right) + \frac{\partial}{\partial y} \left(D_y \frac{\partial}{\partial y} H(x, y, t) \right) + f(I(x, y)) \end{aligned} \quad (2)$$

This 2D formulation captures how pixel information spreads (diffusion) and propagates (advection) across spatial coordinates (x, y) , which is essential for handling blur removal and denoising tasks where spatial relationships are critical.

2) Iterative Solver Embedded in Training

Following the discretization scheme proposed in [19], we adopt a finite-difference approach to make the PDE solver practical for end-to-end training. To solve Eq. 2 numerically, we discretize the spatial and temporal domains with step sizes δx , δy , and δt . The iterative update rule for feature map evolution becomes:

$$\begin{aligned} LH_{x,y}^{k+1} = MH_{x,y}^{k-1} - 2(u_x + v_y) \delta t H_{x,y}^k + 2\delta t f(I(x, y)) \\ + (-A_x + 2B_x) H_{x+1,y}^k + (A_x + 2B_x) H_{x-1,y}^k \\ + (-A_y + 2B_y) H_{x,y+1}^k + (A_y + 2B_y) H_{x,y-1}^k, \end{aligned} \quad (3)$$

where the coefficients are defined following [19] as:

$$\begin{aligned} L &= 1 + 2B_x + 2B_y, & M &= 1 - 2B_x - 2B_y, \\ u_x &= \frac{u_{x+1,y} - u_{x-1,y}}{2\delta x}, & v_y &= \frac{v_{x,y+1} - v_{x,y-1}}{2\delta y}, \\ A_x &= \frac{u \delta t}{2\delta x}, & A_y &= \frac{v \delta t}{2\delta y}, \\ B_x &= \frac{D_x \delta t}{\delta x^2}, & B_y &= \frac{D_y \delta t}{\delta y^2}. \end{aligned} \quad (4)$$

This discretized solver acts as differentiable layers within our restoration network, ensuring that learned features evolve under PDE-guided dynamics. By adapting the methodology from [19] to image restoration, we enable the training procedure to jointly leverage physics-based evolution and data-driven learning, improving the robustness to unseen blurs, including blur and noise. The iterative nature allows for gradual refinement of features through multiple PDE steps, where each iteration applies the physical constraints while preserving gradient flow for end-to-end optimization. When integrated into existing SOTA restoration architectures, this PDE-based global layer improves network performance. This allows us to reduce the complexity of existing NN-based restoration models while keeping the perceptual quality of the image. Specifically, the global layer's ability to propagate information across spatial coordinates through Eq. 3 reduces the dependency on deep feature extraction hierarchies.

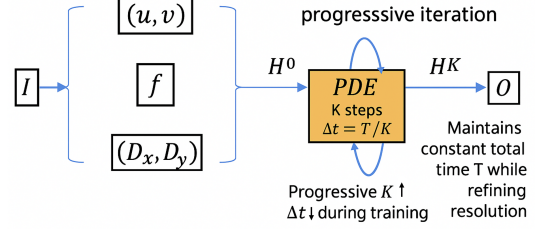


Fig. 2. The Global Layer using PDE models motion blur through learned parameters (u, v) , blur kernel f , and diffusion coefficients (D_x, D_y) . Progressive increase of iteration steps K during training maintains constant total time T while improving computational efficiency, following the diffusion PDE framework of [19].

B. Training Stability and Iteration Constraints

While the PDE-based global layer provides theoretical advantages for spatial information propagation, practical implementation reveals critical stability issues that must be carefully addressed. Our empirical analysis demonstrates that attempting to train directly with multiple PDE iterations ($K > 1$) from the start leads to severe gradient instability. However, this limitation can be overcome through a progressive training strategy that gradually increases iteration count while the network stabilizes.

1) Gradient Instability in Multi-Iteration Schemes

When attempting to train directly with multiple PDE iterations ($K > 1$), we observe severe gradient instability that prevents effective backpropagation. The iterative nature of Eq. (2) creates a computational graph where gradients must traverse through K sequential finite-difference operations, each involving spatial derivatives and coefficient updates. Each iteration involves multiple multiplications with the learned velocity and diffusion parameters, and spatial derivative approximations. Backpropagating gradients through this deep computational chain compounds these operations, leading to unstable gradient magnitudes that prevent effective training. This is particularly problematic because the diffusion terms B_x and B_y in Eq. (4) act as smoothing operators.

2) Progressive Iteration Strategy

To address these challenges, we employ a progressive iteration strategy where K is gradually increased during training from $K = 1$ to $K = 5$. To maintain a constant total integration time T for each layer, we correspondingly reduce Δt by the same factor such that $T = K \cdot \Delta t = \text{constant}$. This enables the use of finer temporal resolution as the network stabilizes during training.

The strategy proceeds in three phases: Phase 1 begins with $K = 1, \Delta t = 1.0$ for the first 10 epochs, establishing a stable training baseline. Phase 2 transitions to $K = 3, \Delta t = 0.333$ for epochs 10–20, introducing moderate temporal refinement. Phase 3 increases to $K = 5, \Delta t = 0.2$ for the remaining training, achieving fine temporal discretization. This gradual increase mitigates gradient instability while enabling the benefits of multi-iteration PDE solving.

As training progresses and gradient fluctuations diminish, the network can handle finer temporal discretization without encountering the vanishing gradient problems that would occur if starting directly with $K = 5$. This approach draws parallels to progressive training methodologies employed in GANs and classical autoencoders, as well as iterative refinement techniques commonly used in numerical optimization. The effectiveness of this progressive schedule is empirically validated in Section IV-C, where we demonstrate that direct training with $K = 5$ fails to converge while progressive training succeeds.

4 EXPERIMENTS

A. Datasets

To evaluate the effectiveness of our proposed PDE-based global layer with progressive training, we conduct comprehensive experiments on three widely adopted benchmarks for image deblurring:

- **GoPro Dataset** [5]: A large-scale synthetic blur dataset comprising 3,214 blurred/sharp image pairs extracted from high-frame-rate videos captured with GoPro cameras. The dataset features diverse dynamic scenes with realistic motion blur patterns, making it a standard benchmark for evaluating motion deblurring algorithms under controlled conditions.
- **RealBlur-R** [12]: A real-world blur dataset containing images captured with camera shake using various handheld devices. This dataset provides authentic blur degradation that better reflects practical imaging scenarios, enabling evaluation of our method’s robustness to natural blur variations encountered in real photography.
- **RealBlur-J** [12]: A complementary real-world dataset focusing on images with motion blur caused by object movement and camera motion in JPEG format. It offers additional validation for our approach across different types of realistic blur patterns and compression artifacts commonly found in real-world applications.

B. Integration into State-of-the-Art Models

We integrate our PDE-informed global layer into several representative state-of-the-art architectures at the network bottleneck, positioned after the encoder and before the decoder at the lowest spatial resolution:

- **FFTformer** [28]: A frequency-domain transformer that exploits global dependencies through Fourier transforms for efficient image restoration. We insert the PDE-based global layer after the frequency domain processing blocks, allowing it to refine the globally-aware features before spatial reconstruction.
- **NAFNet** [11]: A baseline architecture designed for efficiency in image restoration with simplified activation-free residual blocks that achieve competitive performance with reduced computational overhead. The PDE layer is integrated within the middle stages of the encoder-decoder structure, specifically after the deepest feature extraction layers where receptive fields are largest.

- **Restormer** [9]: A transformer-based architecture optimized for image restoration, particularly effective in handling complex degradations through multi-scale feature processing and cross-attention mechanisms. We incorporate the PDE layer between the encoder and decoder sections, enabling global context modeling at the feature bottleneck where spatial resolution is lowest but semantic information is richest.
- **Stripformer** [30]: A transformer-based architecture that employs strip-based attention mechanisms for efficient long-range dependency modeling in image restoration tasks. The PDE-based global layer is positioned at the bottleneck of the architecture, at the lowest spatial resolution where it can most effectively capture long-range dependencies and enhance the strip-based attention’s ability to model directional blur patterns before the reconstruction phase begins.

C. Progressive Training Validation

To empirically validate that gradient instability prevents direct multi-iteration training, we compare two training strategies on NAFNet with 5 PDE layers on GoPro datasets:

TABLE I
PROGRESSIVE TRAINING IS NECESSARY FOR MULTI-ITERATION PDE SCHEMES. HERE ARE THE RESULTS ON THE GOPRO DATASET FOR THE NAFNET MODEL.

Strategy	K	Final PSNR	Converged
Direct	5 (fixed)	-	(diverged epoch 18)
Progressive	1→3→5	34.01	Yes

Direct training with $K=5$ resulted in loss divergence at epoch 18, with gradient norms exceeding 10^3 , confirming the gradient instability. Progressive training successfully enabled stable convergence to $K=5$ by allowing the network to adapt gradually to increasing gradient path lengths.

D. Training Configuration and Progressive Strategy

1) Progressive Training Protocol

Given the stability constraints identified in Section III-B, we implement a progressive iteration strategy during training of the PDE-enhanced architectures. Each model (FFTformer+PDE, NAFNet+PDE, Restormer+PDE, Stripformer+PDE) is trained end-to-end with the integrated PDE global layers using progressive temporal refinement.

The progressive iteration schedule is implemented as follows:

- **Phase 1** (first 10 epochs): $K = 1$, $\Delta t = 1.0$, Total time $T = 1.0$
- **Phase 2** (epochs 10-20): $K = 3$, $\Delta t = 0.333$, Total time $T = 1.0$
- **Phase 3** (epoch 20 till end of training): $K = 5$, $\Delta t = 0.2$, Total time $T = 1.0$

During training, we begin with coarse temporal discretization using larger iteration steps and gradually increase the number of iterations while correspondingly reducing the temporal step size to maintain constant total integration time $T = K \cdot \Delta t$. The PDE parameters (velocity components u , v , and diffusion

TABLE II

QUANTITATIVE COMPARISON OF MODELS WITH AND WITHOUT THE PDE-BASED GLOBAL LAYER ON GoPro, REALBLUR-R AND REALBLUR-J DATASETS. THE NUMBER OF PDE LAYERS FOR OUR MODELS IS SET TO 5 AND $K = 5$.

Model	GoPro		RealBlur-R		RealBlur-J	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FFTformer	34.21	0.969	40.11	0.973	32.62	0.933
FFTformer+PDE	34.45	0.971	40.63	0.977	32.92	0.934
NAFNet	33.71	0.967	39.89	0.973	32.01	0.920
NAFNet+PDE	34.01	0.969	40.10	0.974	32.98	0.932
Restormer	32.92	0.961	40.15	0.974	32.88	0.933
Restormer+PDE	33.77	0.968	40.68	0.976	33.15	0.937
Stripformer	33.08	0.962	39.84	0.974	32.48	0.929
Stripformer+PDE	33.79	0.968	40.89	0.977	33.12	0.936

TABLE III

COMPUTATIONAL COMPLEXITY COMPARISON IN TERMS OF GMACS FOR MODELS WITH AND WITHOUT PDE-BASED GLOBAL LAYER. THE NUMBER OF PDE LAYERS FOR OUR MODELS IS SET TO 5 AND $K = 5$.

Model	GMACS	Model	GMACS
FFTformer	131.45	FFTformer+PDE	132.87
NAFNet	63.64	NAFNet+PDE	64.24
Restormer	141.00	Restormer+PDE	141.60
Stripformer	169.89	Stripformer+PDE	172.31

coefficients D_x, D_y) are initialized using Xavier initialization scaled by 0.01. The iteration count K is doubled while the temporal step Δt is halved at each phase transition, ensuring consistent total integration time across all phases while enabling progressively finer temporal resolution as the network stabilizes during training.

2) Training Hyperparameters

For all experiments, we use the following training configuration:

- **Training hyperparameters:** Optimizer, learning rate schedule, batch size, training epochs, loss function, and data augmentation are adopted from each architecture's original paper
- **PDE-specific parameters:**
 - Spatial discretization: $\delta x = \delta y = 1.0$
 - Temporal step: $\delta t = 1.0$ (iteration constraint)
 - Training epochs: 10 for Phase 1, 10 for Phase 2, same number of iterations as mentioned in the papers for the respective baseline model for Phase 3

E. Evaluation Protocol

We conduct a comprehensive evaluation to assess the impact of our PDE-based global layer on restoration performance and computational efficiency. Each baseline architecture is tested both with and without the integrated PDE layer under identical training and testing conditions.

Performance Metrics:

- **PSNR (Peak Signal-to-Noise Ratio):** Measures pixel-level reconstruction fidelity between restored and ground truth images.

TABLE IV

PERFORMANCE COMPARISON ACROSS DIFFERENT ITERATION COUNTS K FOR THE NAFNET PDE-BASED GLOBAL LAYER MODELS ON GoPro, REALBLUR-R AND REALBLUR-J DATASETS. THE NUMBER OF PDE LAYERS IS SET TO 5.

Iterations (K)	GoPro		RealBlur-R		RealBlur-J	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
$K = 0$	33.71	0.967	39.89	0.973	32.01	0.920
$K = 1$	33.69	0.967	39.94	0.973	32.09	0.920
$K = 3$	33.88	0.968	40.03	0.974	32.67	0.929
$K = 5$	34.01	0.969	40.10	0.974	32.98	0.932
$K = 7$	34.03	0.969	40.15	0.974	33.00	0.932

TABLE V

PERFORMANCE COMPARISON ACROSS DIFFERENT NUMBERS OF PDE-BASED GLOBAL LAYERS ON NAFNET FOR GoPro, REALBLUR-R, AND REALBLUR-J DATASETS. $K = 5$ FOR ALL THE RESULTS PRESENTED IN THIS TABLE.

Number of Layers	GoPro		RealBlur-R		RealBlur-J	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
0 Layer	33.71	0.967	39.89	0.973	32.01	0.920
1 Layer	33.80	0.967	39.94	0.972	32.15	0.923
3 Layers	33.92	0.968	40.02	0.973	32.57	0.927
5 Layers	34.01	0.969	40.10	0.974	32.98	0.932

- **SSIM (Structural Similarity Index):** Evaluates perceptual quality by comparing structural information, luminance, and contrast.

Computational Efficiency Metrics:

- **GMACS (Giga Multiply-Accumulate Operations):** Quantifies computational complexity to assess the overhead introduced by the PDE-based global layers across different architectures.

Evaluation Setup: All models are evaluated on the test splits of GoPro, RealBlur-R, and RealBlur-J datasets using input images of size 256×256 pixels. For computational complexity analysis, we measure GMACS using the final trained models with the complete progressive training protocol ($K = 5$, $\Delta t = 0.2$). This comprehensive evaluation framework enables rigorous analysis of how the proposed PDE-informed global layer contributes to restoration quality and the associated computational overhead across multiple backbone architectures.

This setup allows us to rigorously analyze how the proposed PDE-informed global layer contributes to restoration quality, perceptual improvement, and computational efficiency across multiple backbones.

5 DISCUSSION

A. Consistent Quality Improvements with Minimal Overhead

Our approach achieves consistent improvements across all tested architectures and datasets (Table II), with PSNR gains ranging from +0.24 dB to +1.05 dB and SSIM improvements from +0.001 to +0.007. These improvements are particularly notable given the minimal computational overhead, adding only 0.6 to 2.4 GMACS (approximately 1%) depending on the base architecture.

The universality of improvements across diverse architectures validates the generality of our approach. Both CNN-based (NAFNet) and transformer-based architectures (FFT-

former, Restormer, Stripformer) benefit from PDE integration, demonstrating that physics-informed feature evolution provides complementary advantages regardless of the underlying architectural paradigm. This broad applicability suggests that PDE layers can be readily integrated into existing restoration pipelines without architecture-specific modifications.

Importantly, these quality improvements translate to meaningful perceptual benefits. SSIM gains of 0.001-0.007 based on Table II indicate enhanced structural preservation and texture fidelity, which are critical for downstream vision tasks and human perception. The combined PSNR and SSIM improvements demonstrate that our method enhances both pixel-level accuracy and perceptual quality—a dual benefit that purely data-driven approaches often struggle to achieve simultaneously.

B. Architecture and Dataset-Specific Performance Insights

While all architectures benefit from PDE integration, the magnitude of improvement varies meaningfully across base architectures in Table II, ranging from +0.38 dB average (FFTformer) to +0.71 dB average (Stripformer). This variance provides valuable insight into deployment strategies. FFTformer’s frequency-domain processing already provides strong global receptive fields, so PDE layers offer complementary but smaller gains. Conversely, Stripformer’s strip-based attention benefits substantially from the omnidirectional feature propagation provided by PDE layers, suggesting that architectures with directional biases gain most from our approach.

Dataset-specific performance patterns further demonstrate the method’s strengths. Based on Table II RealBlur-J shows the largest improvements (+0.64 dB average), significantly outpacing gains on GoPro (+0.36 dB average) and RealBlur-R (+0.44 dB average). This pattern reveals a key advantage: PDE-based iterative refinement particularly excels when handling compound degradations. RealBlur-J combines motion blur with JPEG compression artifacts, and the iterative feature evolution appears to disentangle multiple degradation sources more effectively than single-pass convolutions. This suggests our approach is especially valuable for real-world deployment scenarios where images exhibit multiple, interacting degradations.

Tables IV and V reveal optimal configurations that maximize performance while maintaining efficiency. The K iteration ablation shows steady improvements from K=1 to K=5 (+0.30 dB on GoPro), with K=7 providing only marginal additional gains (+0.02 dB). Similarly, the layer count ablation demonstrates that 5 layers achieve near-optimal performance, with diminishing returns beyond this point. These results indicate natural convergence properties rather than arbitrary cutoffs. For deployment, K=3 with 3 layers provides benefits at reduced cost, while K=5 with 5 layers represents the optimal configuration for applications demanding maximum quality. This flexibility enables practitioners to choose appropriate trade-offs for their specific resource constraints and quality requirements.

6 CONCLUSIONS

This paper presents a novel approach to motion deblurring by integrating physics-informed partial differential equation (PDE) dynamics into state-of-the-art restoration architectures

through a progressive training strategy. We use a PDE-based global layer leveraging advection-diffusion equations to model feature motion deblurring, naturally capturing the directional flow characteristics of motion blur while enabling principled global spatial modeling through physics-informed constraints.

Our comprehensive experiments on standard motion deblurring benchmarks (GoPro, RealBlur-R, and RealBlur-J) validate the effectiveness of physics-informed feature processing. The consistent improvements in PSNR (0.2 to 1.0 dB) and SSIM (0.001 to 0.007) metrics across all tested architectures (FFTformer, NAFNet, Restormer, and Stripformer) based on Table II demonstrate that our approach successfully bridges the gap between mathematical modeling and deep learning for image restoration. Importantly, these quality improvements are achieved with minimal computational overhead, adding only approximately 1% to inference GMACs (0.6 to 2.4 GMACs depending on the base architecture) based on Table III.

The progressive training strategy addresses critical stability issues inherent in multi-iteration PDE schemes. By gradually increasing the number of PDE iterations from $K = 1$ to $K = 5$ while correspondingly reducing the temporal step size to maintain constant total integration time, our approach ensures stable gradient flow and convergence throughout the training process. This strategy enables successful integration of iterative PDE solvers into existing restoration architectures without encountering the gradient vanishing problems that would otherwise prevent effective backpropagation.

The results demonstrate that incorporating mathematical physics principles into deep learning architectures can enhance restoration quality while maintaining computational efficiency comparable to baseline methods. This physics-informed approach establishes a promising direction for leveraging domain knowledge through PDE-based modeling in image restoration tasks. By treating blur removal as a guided feature evolution process constrained by advection-diffusion dynamics, our method achieves more effective restoration across diverse blur scenarios.

REFERENCES

- [1] L. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, 1974.
- [2] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55-59, 1972.
- [3] S. Cho and S. Lee, "Fast motion deblurring," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 1-8, 2009.
- [4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 787-794, 2006.
- [5] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3883-3891.
- [6] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8174-8182.
- [7] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurgAN: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183-8192.

- [8] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5978-5986.
- [9] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14821-14831.
- [10] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 1833-1844.
- [11] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European Conference on Computer Vision*, 2022, pp. 17-33.
- [12] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 184-201.
- [13] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 168-186, 2012.
- [14] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 769-777.
- [15] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2017.
- [17] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision Workshops*, 2018, pp. 63-79.
- [18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, 2017.
- [19] A. Kag, Z. Zhang, and V. Saligrama, "Condensing CNNs with partial differential equations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 895-904.
- [20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [21] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105-6114.
- [22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [23] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 10012-10022.
- [24] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17683-17693.
- [25] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12299-12310.
- [26] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," in *Advances in Neural Information Processing Systems*, 2021, pp. 15908-15919.
- [27] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 568-578.
- [28] L. Kong, J. Dong, J. Ge, M. Li, and J. Pan, "Efficient frequency domain-based transformers for high-quality image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5886-5895.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.
- [30] F. Tseng, Y. Luo, Y. Zhang, S. Xu, Y. Wu, and Y. Tai, "Stripformer: Strip transformer for fast image deblurring," in *Proceedings of the European Conference on Computer Vision*, 2022, pp. 146-162.