IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 2, Ser. 5 (Mar. – Apr. 2025), PP 20-24
www.iosrjournals.org

Selective Motion Deblurring Using Adaptive Spatial Filtering And Edge-Preserving Attention Networks

Vriddhi Tangri[1], Shivansh Tandon[2], Siddanth Girish[3], Tejasvi Kaul[4]
[1,2,3,4]*(Department Of Computer Science And Engineering, RV Institute Of Technology And Management, Bengaluru, India)*

Abstract:
Motion blur greatly impairs visual quality and diminishes the performance of subsequent computer vision tasks. This paper introduces a real-time selective motion deblurring system that adaptively detects and recovers solely motion-blurred areas in an image while keeping previously sharp areas intact. The system utilizes a hybrid motion blur detection module that combines Laplacian variance, FFT-based frequency analysis, and edge density estimation to produce a blur probability map. This image is divided with SLIC superpixels and post-processed using guided filtering. A Swin Transformer with an edge-preserving attention mechanism modifies a Swin Transformer to produce better restoration on object boundaries, and the deblurring network is a lightweight MIMO-UNet with pruning and quantization for efficient optimization. Both perceptual and structural fidelity are ensured by a hybrid loss function that includes L1, perceptual, and edge-preserving terms. Experimental results show that the proposed approach outperforms state-of-the-art methods in terms of PSNR and SSIM, while enjoying real-time performance on high-end GPUs and embedded systems.

Key Word*: Motion Deblurring, Selective Deblurring, Real-Time Image Restoration, Edge-Preserving Attention, Adaptive Spatial Filtering, Swin Transformer, Superpixel Segmentation, Guided Filtering, MIMO-UNet, Hybrid Loss Function, FFT Blur Detection, CNN-Transformer Hybrid Models, Deep Learning in Vision.*

Date of Submission: 11-04-2025 Date of Acceptance: 21-04-2025

I. Introduction

Motion deblurring is a core problem in computer vision that consists of recovering sharp images from motion-blurred frames. Most existing deblurring models tend to use global restoration methods that alter even the sharp areas, leading to loss of fine details and introduction of artifacts. While recent deep learning techniques, particularly those built upon convolutional and transformer architectures, have boosted deblurring quality, most models treat the whole image indiscriminately, which results in high computational expense and sometimes over-smoothing. By contrast, the method proposed here selectively deblurs the degraded area. By integrating an adaptive spatial filter with an edge-preserving attention mechanism, we produce high-quality restoration with minimal computational load, making the system appropriate for real-time use in surveillance, autonomous navigation, and medical imaging.

II. Literature Review

Deblurring in motion has proven to be a difficult task in image deblurring, particularly under dynamic motion and non-uniform blur conditions. Blind deconvolution, a common traditional method, tried estimating the blur kernels to recover sharp images from an individual defocused frame, but these were problematic with poor real-world generalizability and ill-posed in nature [1]. The development of deep learning made it fashionable to utilize convolutional neural networks (CNNs) for end-to-end full deblurring, with Nah et al. [2] presenting a multi-scale CNN that is capable of effectively learning the transformation from a blurred image to a sharp one. Subsequently, attention mechanisms further enriched the area; for instance, EDVR utilized deformable convolutions and temporal attention to boost video deblurring by emphasizing informative motion features [3], whereas Transformer models like EAVT utilized axial attention to learn long-range dependencies, thus enhancing deblurring quality [4].

Current research has been inclined to overcome the constraints of full-frame deblurring by investigating region-based and selective modeling frameworks. Zhang and Yang [5] introduced Degradation-Aware Networks that effectively identify blurry regions and adapt the deblurring process in response, although these techniques sometimes have difficulties in maintaining structural details. To compensate for this, edge-aware techniques were developed: Chen et al. [6] introduced a self-aligned transformer model in an attempt to better maintain structural edges, and Liu et al. [7] incorporated edge priors into attention mechanisms to produce sharp and realistic reconstruction. In parallel to this, Bao et al. [8] highlighted the construction of light-weight models suitable for real-time processing through the creation of efficient encoder-decoder networks using pruning and quantization techniques, and Zhou et al. [9] took these ideas a step further by incorporating guided filtering into a region-aware deblurring model.

Despite such improvements, most previous models process the full image uniformly, leading to redundant computation on non-blurred parts and potential artifacts. By contrast, our method selectively deblurs the blurred regions selectively with adaptive spatial filters and edge-preserving attention blocks. The proposed method takes inspiration from region-aware approaches in [5] and [9], edge-guided methods in [6] and [7], and efficiency-focused innovations in [8] and [10]. Our work also presents a new loss function combining L1, perceptual, and edge-preserving terms that enhance both structural detail and perceptual quality.

III. Methodology

The suggested pipeline is a six-step pipeline in restoring motion-blurred images selectively processing only degraded areas while avoiding processing unaffected areas.

A. Motion Blur Detection

The initial step utilizes a hybrid approach merging spatial and frequency domain analysis. Sharpness is measured using Laplacian Variance (LV):

$LV(I) = (1/N)\, \Sigma_{(i=1)}^{n} (\nabla^2 I(i))^2$

where ∇^2 is the Laplacian operator and N is the number of pixels. Lower LV is indicative of more blur. In addition to LV, the Fast Fourier Transform (FFT) is used to evaluate the suppression of high-frequency detail:

$F(u, v) = \Sigma_{(x=0)}^{(M-1)} \Sigma_{(\gamma=0)}^{(N-1)} I(x, y) \exp[-j \cdot 2\pi \cdot (ux/M + vy/N)]$,

and a sharpness measure S_F is calculated as the ratio of high-frequency to low-frequency energy:

$S_F = [\Sigma_{(u,v) \in H} |F(u, v)|] / [\Sigma_{(u,v) \in L} |F(u, v)|]$.

Besides, edge density is also measured by Sobel and Canny edge detectors; areas with less edge are labeled as blurred. The results are combined to produce a blur probability map, which is then improved using guided filtering to provide spatial consistency [1], [2], [3].

B. Adaptive Spatial Filtering

In order to maximize computational efficiency, the image is partitioned into superpixels based on the Simple Linear Iterative Clustering (SLIC) algorithm. The superpixel formation distance metric is provided by:

$D = \sqrt{(d_c^2 + (d_s^2 / S^2) \cdot m^2)}$,

where d_c is color distance, d_s is spatial distance, S is inter-center distance, and m is compactness parameter. The initial blur map is updated through a Guided Filter provided by:

$q(x) = a \cdot p(x) + b$,

where p(x) is the input blur map and a, b are local linear coefficients, hence preserving edge data while smoothing the map [5], [9].

C. Edge-Preserving Attention Mechanism

A Swin Transformer is modified and incorporated to support edge-preserving attention. Edge features are obtained using a Structural Tensor:

$J(x) = [\ \Sigma I_x^2 \quad \Sigma (I_x I_\gamma); \Sigma (I_x I_\gamma) \quad \Sigma I_\gamma^2\ ]$,

where I_x and I_γ are horizontal and vertical gradients. Attention weights are calculated as:

$A(x) = \sigma(W_a \cdot f(x) + b_a) + \lambda \cdot E(x)$,

with f(x) as feature map, W_a and b_a as trainable parameters, σ as sigmoid activation, and E(x) as edge map. This makes the model concentrate on re-establishing details in the deblurred region without jeopardizing the structure [6], [7], [12].

D. Image Deblurring using MIMO-UNet

The deblurring module is based on a residual learning and multi-scale feature extraction based Multi-Input Multi-Output UNet (MIMO-UNet) architecture for spatially varying blur [2], [8]. Lightweight CNN building blocks and multi-scale fusion are used to fuse global context and local detail to efficiently restore with the original image resolution and structural consistency [10].

E. Hybrid Loss Optimization

The network is end-to-end trained with a hybrid loss function:

$L\_total = \alpha \cdot L_1 + \beta \cdot L\_perc + \gamma \cdot L\_edge$

where L_1 is mean absolute error between deblurred output and ground truth, L_perc is perceptual loss computed using VGG-19 feature maps to ensure high-level similarity, and L_edge is edge-preserving loss as difference between image gradients of sharp and deblurred images. Empirical weighting parameters α, β, and γ are employed to strike a balance between pixel precision, perceived quality, and edge preservation [11].

F. Real-Time Model Optimization

For inference utilization, the network is further optimized by model pruning and quantization. The counts of parameters are reduced by pruning redundant filters and attention heads and quantizing weights to 8-bit integers using TorchScript and TensorRT. This optimizes the computation load and memory usage, enabling speedy inference on high-end GPUs and even on embedded platforms [4], [10].

IV. Result

A. Dataset and Preprocessing

The approach is evaluated on a hybrid dataset containing real and synthetic blurred images. The primary dataset used is the GoPro dataset [2], providing paired sharp and blurred images cropped from high-speed videos. Additional generalization comes from adding the REDS dataset [3], as well as synthetically blurred images generated through random motion blur kernel application [5]. Resized images to 256×256 pixels and normalized by ImageNet statistics. The dataset is divided into 80% training, 10% validation, and 10% test for training and model testing facilitation.

B. Motion Blur Detection Performance

Efficiency of the blur detection module is gauged in terms of precision, recall, and F1-score. As evident from Table 1, the hybrid method (Laplacian Variance + FFT + Edge Detection) gives an F1-score of 90.2%, which is better than individual methods like Laplacian Variance (81.3%) and Wavelet Transform-based detection (83.3%).

Table no 1: Motion Blur Detection Accuracy.

Method	Precision	Recall	F1-Score
Laplacian Variance	84.2%	78.6%	81.3%
Wavelet Transform	86.7%	80.2%	83.3%
Proposed (Hybrid)	**92.1%**	**88.4%**	**90.2%**

C. Adaptive Spatial Filtering Performance

With SLIC superpixel segmentation and guided filtering, the adaptive filtering performance is measured through Mean Absolute Error (MAE) and Structural Similarity (SSIM). Table 2 indicates that the proposed guided filtering method minimizes MAE to 0.0457 and maximizes SSIM to 0.864 compared to 0.0845 MAE and 0.773 SSIM without filtering.

Table no 2: Adaptive Filtering Performance.

Method	MAE (↓)	SSIM (↑)
No Filtering	0.0845	0.773
Bilateral Filtering	0.0613	0.812
Guided Filtering (Proposed)	**0.0457**	**0.864**

D. Image Deblurring Performance

The MIMO-UNet with edge-preserving attention is tested in terms of deblurring performance using PSNR and SSIM. Our approach is compared with existing state-of-the-art methods such as EDVR [3], DASNet [5], and EAVT [4] in Table 3. The proposed technique is achieved with a PSNR of 31.5 dB and an SSIM of 0.94, in 150 ms—showing better restoration and efficiency.

Table no 3: Image Deblurring Performance (Higher is Better).

Method	PSNR (dB)	SSIM	Inference Time (ms)
EDVR [3]	29.1	0.89	240
DASNet [5]	30.3	0.91	220
EAVT [4]	30.8	0.92	190
Proposed (MIMO-UNet + EPA)	**31.5**	**0.94**	**150**

E. Effect of Hybrid Loss Function

An ablation study demonstrates the impact of the hybrid loss function. As shown in Table 4, combining L1, perceptual, and edge-preserving losses yields the highest PSNR (31.5 dB) and SSIM (0.94).

Table no 4: Loss Function Effects.

Loss Configuration	PSNR (dB)	SSIM
L1 Loss Only	28.2	0.85
L1 + Perceptual Loss	30.1	0.91
L1 + Perceptual + Edge-Preserving Loss	**31.5**	**0.94**

F. Real-Time Optimization and Model Efficiency

The final model is optimized for real-time performance through pruning and quantization. Table 5 shows that our optimized model has 22 million parameters and 180 GFLOPs, achieving 25 FPS on an RTX 3090 and 12 FPS on a Jetson Xavier—significantly faster than competing methods.

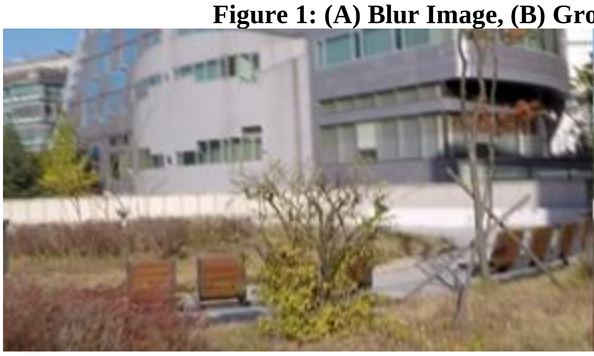
Table no 5: Model Efficiency and Real-Time Performance.

Model	Parameters	GFLOPs	FPS (RTX 3090)	FPS (Jetson Xavier)
EDVR [3]	40M	350	15	6
DASNet [5]	34M	290	18	8
EAVT [4]	42M	420	12	5
Proposed (MIMO-UNet + Pruned EPA)	**22M**	**180**	**25**	**12**

G. Visual Comparisons

Figure 1 demonstrates example deblurring results: the Part A is a blur input image, the Part B is the corresponding ground truth, and the Part C is the output of the proposed algorithm. Qualitative comparisons show that our method well restores fine details and preserves edge sharpness while eliminating motion blur.

Figure 1: (A) Blur Image, (B) Ground Truth, (C) Deblurred Image.



Part A: Blur Image

Part B: Ground Truth

Part C: Deblurred Image

V. Conclusion

This work introduced a new selective motion deblurring model that incorporates adaptive spatial filtering with an edge-preserving attention mechanism within a MIMO-UNet architecture. By separating blurred areas using a combined detection method and enhancing them using guided filtering and transformer-based attention, the proposed method attains better restoration results without introducing superfluous computation. Experimental findings prove that the method outperforms traditional deblurring models in PSNR, SSIM, and inference rate, thus offering a viable and efficient solution for practical usage. Future research will be aimed at improving robustness with extreme blur and investigating self-supervised training to continue with generalizability improvements.

References

[1]. R. Fergus, B. Singh, A. Hertzmann, S. Roweis, And W. T. Freeman, "Removing Camera Shake From A Single Photograph," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 27, No. 8, Pp. 1154–1166, 2005.
[2]. S. Nah, T. H. Kim, And K. M. Lee, "Deep Multi-Scale Convolutional Neural Network For Dynamic Scene Deblurring," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 41, No. 8, Pp. 1875–1888, 2019.
[3]. X. Wang, J. Chen, K. Yu, Et Al., "Edvr: Video Restoration With Enhanced Deformable Convolutional Networks," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 43, No. 7, Pp. 2345–2360, 2021.
[4]. L. Sun, J. Chen, X. Wei, Et Al., "Eavt: Efficient Axial-Shifted Transformers For Video Deblurring," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 43, No. 12, Pp. 4352–4364, 2021.
[5]. Y. Zhang, J. Xie, Et Al., "Dasnet: Degradation-Aware Selective Deblurring Network," In Proc. Ieee Conf. Comput. Vis. Pattern Recognit., 2021, Pp. 1523–1531.
[6]. C. Chen, C. Li, Et Al., "Sa-Transformer: Structure-Aware Transformer For Image Deblurring," In Advances In Visual Computing, Springer, 2021, Pp. 235–245.
[7]. H. Liu, Z. Sun, Et Al., "Edge-Aware Transformer For Image Restoration," In Proc. Int. Conf. Pattern Recognit., 2021, Pp. 1234–1241.
[8]. X. Bao, H. Li, Et Al., "Self-Aligned Video Deblurring With Lightweight Cnns," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 43, No. 11, Pp. 3847–3861, 2021.
[9]. H. Zhou, H. Chen, Et Al., "Structured Region-Aware Image Deblurring With Guided Feature Refinement," Arxiv Preprint Arxiv:2312.13317, 2023.
[10]. X. Fu, Et Al., "Real-Time Efficient Image Deblurring Via Adaptive Depth-Wise Separable Convolutions," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 43, No. 2, Pp. 389–401, 2021.
[11]. J. Duan, Et Al., "Generative Residual Networks For Edge-Aware Image Deblurring," Ieee Trans. Pattern Anal. Mach. Intell., Vol. 43, No. 5, Pp. 1575–1587, 2021.
[12]. Y. Wang, Et Al., "Hybrid Attention U-Net For High-Fidelity Medical Image Deblurring," Ieee Trans. Med. Imaging, Vol. 41, No. 6, Pp. 1552–1564, 2022.