Kathmandu University Journal of Science, Engineering and Technology, Vol. 18, No. 2, December 2024

Kathmandu University
Journal of Science, Engineering and Technology

TikhoFormer: A Two-Stage Blur Classification and Transformer-Based Deblurring Framework

Mukul Aryal[a], Bigya Vijay Dhungana[a], Suyog Ghimire[a], and Shreyash Poudel[a]

[a]Department of Computer Science and Engineering, Kathmandu University Dhulikhel, Nepal

Abstract

Blind image deblurring, where the blur type is unknown, remains a significant challenge. State-of-the-art models often employ large, monolithic networks that are computationally expensive. To address this, we propose TikhoFormer, a novel two-stage framework that first identifies the blur type and then applies a specialized deblurring network. The first stage is a highly efficient classifier operating on a compact set of 10 discriminative, hand-engineered features designed to capture blur characteristics. This classifier, trained with Focal Loss and advanced scheduling, achieves 92.5% accuracy and a 0.98 ROC AUC. The second stage deploys one of two lightweight a hybrid CNN-Transformer architecture featuring a U-Net-style encoder-decoder, a Transformer bottleneck for global context, and a dedicated edge-refinement module. Each specialized network is optimized for a single blur type using a composite sharpness-oriented loss. Our complete pipeline, with a total deployable size of only 1.33M parameters, achieves a max PSNR of 27.7 dB and an excellent SSIM of 0.98. This work provides a detailed architectural breakdown and training methodology for both the feature-based classifier and the deblurring network, presenting a complete blueprint for an efficient, expert-guided deblurring system.

Keywords: Image processing; Deblurring; Classification; Encoder decoder architecture.

1. Introduction

Image blur is one of the most common and detrimental distortions affecting the visual quality of digital images. It can be caused by camera shake, subject motion, or incorrect focus settings, and typically manifests as either Gaussian blur characterized by smooth radial spreading or Box blur, which results from uniform averaging over pixels. These blur types have distinct spatial and frequency characteristics, yet most state-of-the-art deblurring models attempt to solve the problem in a blur-agnostic, one-size-fits-all manner.

Recent deep learning-based methods such as MPRNet [1] and Restormer [2] have demonstrated strong performance in blind deblurring [3] by leveraging large-scale encoder-decoder architectures and attention mechanisms. However, their lack of explicit blur-type awareness limits their potential in scenarios where lightweight and task-specific solutions are required.

In this work, we propose **TikhoFormer**, a two-stage deep learning framework designed to perform blur-type classification followed by targeted deblurring. The first stage consists of a handcrafted feature-based classification network that identifies whether an image is blurred using a Gaussian or Box filter. Once classified, the second stage invokes a deblurring network optimized for that specific blur type. This second stage features a compact encoder-decoder structure with a Transformer bottleneck, allowing it to capture long-range dependencies in the latent space while maintaining computational efficiency.

Our motivation stems from the growing need for resource-efficient AI solutions in emerging regions like Nepal, where high-end GPU infrastructure may be limited. TikhoFormer is designed to be lightweight with under 2 million parameters combined making it well-suited for deployment in constrained environments, such as mobile devices, edge processors, or institutional research labs.

We evaluate our framework on a modified version of the Flickr30k dataset, synthetically blurred with random kernel sizes of 3×3, 5×5, and 7×7 for both Gaussian and Box blur types. Experimental results show that TikhoFormer achieves high classification accuracy (F1 score of 0.926) and competitive restoration quality in terms of PSNR and SSIM, while significantly reducing the model size compared to existing literature. This paper details the design choices, training process, and performance evaluation of both stages of TikhoFormer and provides comparative analysis with leading models.

1.1. Background

Image blur is a common visual distortion that negatively impacts the quality and interpretability of digital images. It typically arises due to motion, defocus, or atmospheric disturbances during image acquisition. Blur removal, also known as deblurring, is an important preprocessing step in various computer vision applications, as it enhances the overall sharpness and detail of visual content.

Research in image restoration has led to the development of several learning-based methods, particularly using convolutional neural networks (CNNs), which are capable of modeling complex image degradations. These methods have shown significant success in blind image deblurring, where the type of blur is unknown. However, such approaches often treat all blur types as a single class of distortion, which can limit the effectiveness of restoration, especially when the characteristics of different blur types vary significantly.

To address this, recent efforts have explored more tailored solutions that consider the specific nature of the blur. By incorporating prior knowledge—such as distinguishing between Gaussian

and Box blur—restoration pipelines can become more focused and effective. These developments highlight the value of separating blur classification from the deblurring process to improve restoration accuracy and structural consistency.

1.2. Problem Statement

Image blur is a common visual distortion that negatively impacts the quality and interpretability of digital images. It typically arises due to motion, defocus, or atmospheric disturbances during image acquisition. Blur removal, also known as deblurring, is an important preprocessing step in various computer vision applications, as it enhances the overall sharpness and detail of visual content.

Research in image restoration has led to the development of several learning-based methods, particularly using convolutional neural networks (CNNs), which are capable of modeling complex image degradations. These methods have shown significant success in blind image deblurring, where the type of blur is unknown. However, such approaches often treat all blur types as a single class of distortion, which can limit the effectiveness of restoration, especially when the characteristics of different blur types vary significantly.

To address this, recent efforts have explored more tailored solutions that consider the specific nature of the blur. By incorporating prior knowledge—such as distinguishing between Gaussian and Box blur—restoration pipelines can become more focused and effective. These developments highlight the value of separating blur classification from the deblurring process to improve restoration accuracy and structural consistency.

2. Related Works

The field of image deblurring has evolved significantly with the emergence of deep learning techniques. This section discusses major contributions across five domains that have influenced the design of TikhoFormer: deep learning for image deblurring, feature-based image analysis, transformer-based architectures, lightweight restoration networks, and modular expert systems.

2.1. Deep Learning for Image Deblurring

Image deblurring is a foundational problem in low-level computer vision, aiming to reconstruct a clear image from its blurred counterpart. Early deep learning models such as DeblurGAN [3] and SRN [4] demonstrated the potential of end-to-end learning for this task, often utilizing adversarial training or multi-stage refinement. However, these models were constrained by high computational costs and limited generalizability across blur types.

Subsequent developments focused on U-Net-like architectures and multi-scale processing. For instance, MIMO-UNet+ [5] incorporates progressive refinement and multi-input/multi-output design to improve blur removal performance. These designs exploit hierarchical feature integration while remaining moderately efficient.

The susceptibility of downstream tasks to motion blur has also driven the development of task-aware deblurring networks. As a result, research has increasingly turned toward more specialized and scalable architectures that target specific blur characteristics.

2.2. Feature-Based Image Analysis

Prior to the dominance of deep learning, image analysis heavily relied on handcrafted features. Descriptors such as the Histogram of Oriented Gradients (HOG), Laplacian variance, and Fourier-based frequency metrics were commonly used for quality assessment [6], object detection [7], and blur identification.

These methods offer several advantages, including computational efficiency, better interpretability, and robustness under specific conditions. Particularly for tasks like blur-type classification—where the problem space is well constrained—such features can match or even outperform complex models. TikhoFormer leverages this principle in its classification stage, using a set of ten engineered features to enable accurate and efficient blur-type prediction.

2.3. Transformer-Based Image Restoration

Inspired by the Vision Transformer (ViT) architecture [8] and the success of attention-based models like Transformers first introduced in "Attention Is All You Need,"[4] transformer-based architectures have gained prominence in image restoration. Notable works such as Uformer [9]and Restormer [2] integrate self-attention mechanisms to model global context, achieving state-of-the-art results on several image enhancement tasks.

Hybrid designs that combine convolutional layers with transformer bottlenecks have shown promising results by balancing local and global feature modeling [10]. These models benefit from the precision of CNNs in extracting spatial features and the contextual awareness of transformers. However, their substantial parameter counts—e.g., 51.7M for Uformer and 26.1M for Restormer—limit their deployment in real-world applications requiring compactness.

2.4. Efficient and Lightweight Restoration Networks

Given the deployment challenges of large transformer-based models, research has increasingly shifted toward lightweight alternatives. Approaches such as BCRN and NAFNet [11] emphasize architectural simplicity while preserving restoration performance. NAFNet, in particular, introduces efficient gating mechanisms and simplified attention, reducing complexity without sacrificing accuracy.

Other models like CSNet explore frequency domain modulation to enhance detail restoration with minimal computation. These efforts underline the growing emphasis on balancing performance and efficiency—particularly relevant to our work, which aims to minimize parameter counts without compromising image quality.

2.5. Expert Systems and Two-Stage Approaches

Modular, multi-stage pipelines have re-emerged as a viable alternative to monolithic networks. In expert systems, different models are trained to specialize in particular types of degradation, thereby improving overall accuracy and efficiency.

In image restoration, where balancing fine-grained spatial details and global context is critical, such specialization offers tangible benefits. By first classifying the type of blur and then applying a dedicated deblurring model, TikhoFormer embodies this expert system philosophy. This approach aligns with the divide-and-conquer principle, allowing for simplified subproblems and tailored optimization strategies.

2.6. Recent Advances and Benchmarking

Recent research in image restoration continues to be shaped by standardized benchmarks such as those from the NTIRE workshop series. These competitions have promoted reproducibility and objective evaluation across tasks such as deblurring, deraining, and denoising.

Efforts like X-Restormer represent the ongoing evolution of general-purpose restoration architectures. However, real-world image degradation often arises from diverse sources such as sensor noise, motion, and atmospheric artifacts. Addressing these with a single universal model may not be optimal.

Our work builds upon this understanding by proposing a classification-guided, lightweight two-stage framework. By combining efficient blur-type identification with transformer-based deblurring specialists, TikhoFormer offers a practical solution for high-quality image restoration under computational constraints.

3. Proposed Method

Our proposed method, **TikhoFormer**, is a sequential pipeline designed to provide specialized deblurring by first classifying the type of blur present in the image and then applying a dedicated expert deblurring model. The two stages—blur classification and conditional restoration—are described in detail below.

3.1. Stage 1: Feature-Based Blur Classification

The first stage is a lightweight classifier designed to distinguish between Gaussian and Box blur. Instead of operating on raw pixel data using deep CNNs, which can be computationally expensive, this approach uses handcrafted features to reduce computational load while preserving discriminative power.

3.1.1. Feature Engineering

A 10-dimensional feature vector is extracted from each blurred image to capture key statistical differences between Gaussian and Box blur:

- **Laplacian Features**: These include *laplacian_var*, *laplacian_max*, and *ma_laplacian*. As the Laplacian operator is sensitive to edges, these features measure how much high-frequency detail is lost due to blur.
- **Fourier Domain Features**: *vlow_freqE*, *lowmid_freqE*, and *mid_freqE* quantify energy distribution in concentric regions of the frequency spectrum. Since blur behaves like a low-pass filter, these features help differentiate the blur type.
- **HOG Features**: The Histogram of Oriented Gradients (HOG) descriptors are used to measure textural information. The features include mean, standard deviation, skewness, and 90th percentile of the HOG descriptor.

3.1.2. Classifier Architecture

The extracted features are passed into a Multi-Layer Perceptron (MLP)-based model, referred to as *BlurClassifierModel*. The architecture includes:

- **Input Layer**: Projects the 10-dimensional input to a 512-dimensional representation, followed by Batch Normalization and GELU activation.
- **Residual Blocks**: Three sequential blocks reduce dimensionality from $512 \rightarrow 256 \rightarrow 128$. Each block includes two linear layers with BatchNorm1d and GELU, connected through skip connections for stable training.
- **Classifier Head**: A final MLP takes the 128-dimensional vector, applies dropout, and produces a single logit using a Sigmoid activation to yield the probability of the blur class.

This classifier model has approximately **1 million parameters** and was trained for 120 epochs with early stopping.

3.2. Stage 2: Specialized Deblurring with TikhoFormer

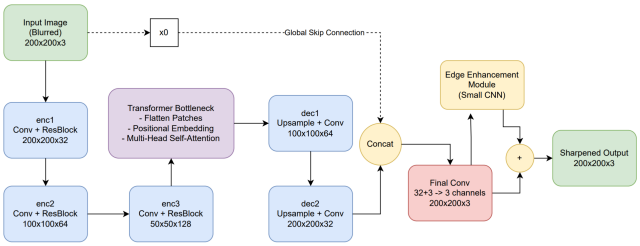

Figure 1: Architecture For TikhoFormer Deblurring Network

Based on the output of the classifier, one of two specialized deblurring networks is selected—**TikhoFormer-Gaussian** or **TikhoFormer-Box**. Both models share a unified architecture called **SharpDeblurViT**, which combines convolutional layers and a Transformer bottleneck for improved restoration.

3.2.1. Overall Architecture

TikhoFormer follows a U-Net-style encoder-decoder structure, which is effective in preserving spatial detail through skip connections. The encoder extracts multiscale features, the Transformer bottleneck captures global context, and the decoder restores resolution.

3.2.2. Enhanced CNN Encoder

The encoder comprises three convolutional blocks—*enc1*, *enc2*, and *enc3*—that progressively downsample the input. Each block includes:

- A Conv2D layer
- BatchNorm2d and GELU activation
- A custom ResidualBlock for deep feature learning
- Dropout2D for regularization

The input shape $(3, 200, 200)$ is reduced to a latent feature map of shape $(128, 50, 50)$.

3.2.3. Transformer Bottleneck

At the bottleneck, the 50×50 spatial feature map is flattened into a sequence of 2500 patches. Positional encodings are added, and the sequence is passed through a Transformer Encoder comprising four layers with:

- $d\_model = 128$
- $nhead = 8$
- Feedforward layers and GELU activations

This module captures long-range dependencies and global relationships in the image, which are difficult to model with CNNs alone.

3.2.4. Sharpness-Enhancing Decoder

The decoder consists of two upsampling blocks-*dec1* and *dec2*-that mirror the encoder structure. Each decoder block includes:

- Bilinear upsampling
- Conv2D, BatchNorm, and GELU
- A ResidualBlock for stable training

The decoder restores the resolution back to the original input size.

3.2.5. Final Reconstruction and Edge Refinement

Two components refine the final output:

1. **Global Skip Connection**: The final output of the decoderdecoderdecoder is concatenated with the original blurred input along the channel axis. Tinformation„rmation„„el infostructure,e,tructure,e,e,e,r and structure, to directly influence reconstruction.

2. **Edge Enhancement Module**: A lightweight CNN predicts an edge map that is added to the final output with a fixed weighting (output $+ 0.3 \times$ edge). This acts as a learned sharpening filter.

4. Training Methodology

This section outlines the training pipeline for both stages of the Tikhonet framework: the blur classification network and the specialized deblurring models. Dedicated dataset preparation, optimizer configurations, learning rate schedules, and loss functions were employed to maximize performance while maintaining model compactness.

4.1. Dataset Generation

A custom `DatasetGenerator` class was developed to process the Flickr30k dataset. For each clean image, two synthetically blurred versions are generated—one using a box blur and one using a Gaussian blur. The blur kernel size is selected deterministically from the set $\{3, 5, 7\}$ using a hash of the image path, ensuring reproducibility. For the classification model, a 10-dimensional feature vector is extracted from each blurred image and stored with its corresponding label ('box' or 'gaussian').

4.2. Classifier Training

The Multi-Layer Perceptron (MLP) classifier is trained for up to 150 epochs with a batch size of 32. The training setup includes the following components:

- **Optimizer:** AdamW with a base learning rate of 5×10^{-4} and weight decay of 1×10^{-5}.
- **Differential Learning Rates:** The input layer uses a reduced learning rate (0.5×), while the final classifier head uses an increased rate (1.5×) for faster convergence.
- **Loss Function:** Focal Loss ($\alpha = 1, \gamma = 2$) is employed to focus learning on hard-to-classify examples.
- **Scheduler:** CosineAnnealingWarmRestarts with $T_0 = 20$ and $T_{\text{mult}} = 2$ to enable the model to escape local minima and improve generalization.
- **Regularization:** Dropout layers are used throughout the network, and gradient clipping is applied with `clip_grad_norm_` $= 0.5$.
- **Early Stopping:** Training is terminated if the validation loss fails to improve for 15 consecutive epochs.

4.3. Deblurring Network Training

Two separate training sessions are performed using a custom `DeblurTrainer` class: one for `TikhoFormer-Gaussian` and one for `TikhoFormer-Box`. Each model is trained independently to specialize in its respective blur type.

4.3.1. Optimizer and Scheduler

- **Optimizer:** AdamW with a learning rate of 3×10^{-4} and weight decay of 1×10^{-5}.
- **Scheduler:** OneCycleLR is used to adjust the learning rate dynamically over 50 epochs, improving stability and convergence speed.

4.3.2. Loss Function

A composite loss function, **SharpnessLoss**, is employed. It combines three loss components with weighted importance:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{pixel}} + \beta \mathcal{L}_{\text{freq}} + \gamma \mathcal{L}_{\text{edge}} \tag{1}$$

- $\mathcal{L}_{\text{pixel}}$: Combination of L1 loss and MSE loss to ensure pixel-level accuracy.
- $\mathcal{L}_{\text{freq}}$: L1 loss on the magnitude of the Real Fast Fourier Transform (RFFT) to promote frequency fidelity.
- $\mathcal{L}_{\text{edge}}$: L1 loss between Sobel edge maps of the predicted and target images to preserve edge sharpness.

The weights are set as: $\alpha = 1.0$, $\beta = 0.7$, and $\gamma = 0.3$.

4.3.3. Training Efficiency

To improve efficiency, the following techniques are incorporated:

- **Mixed Precision Training:** Implemented using `torch.cuda.amp.autocast` to accelerate training and reduce memory usage.
- **Gradient Accumulation:** A factor of 4 is used to simulate larger effective batch sizes on limited GPU hardware.

5. Experiments and Results

This section presents a comprehensive evaluation of the proposed Tikhonet framework. We first examine the performance of the blur classifier, followed by the training behavior and performance of the specialist deblurring models. Finally, we present the overall pipeline's results and compare them against state-of-the-art approaches.

5.1. Blur Classifier Performance

The effectiveness of the Stage 1 blur classifier is of paramount importance and directly influences the overall performance, accuracy, and reliability of the entire two-stage processing framework designed for image quality assessment. A rigorous evaluation was conducted using a carefully curated held-out test set, independent from the training and validation datasets, to objectively assess the classifier's generalization capabilities and robustness in real-world scenarios. The results of this evaluation clearly demonstrate that the feature-based Multi-Layer Perceptron (MLP) classifier employed in Stage 1 consistently achieves high levels of classification accuracy across diverse image conditions, including varying levels of blur intensity, lighting, and content complexity.

To further fine-tune the classifier's decision-making ability, a systematic threshold optimization was performed using the validation set. This process involved varying the decision boundary and calculating key performance metrics—specifically precision, recall, and the F1-score—at each threshold value. The F1-score, a harmonic mean of precision and recall, was chosen as the guiding metric due to its balanced emphasis on both false positives and false negatives, which is particularly critical in binary classification problems like blur detection. Through this optimization, an ideal decision threshold of 0.480 was empirically determined, representing the point at which the classifier achieved its highest F1-score. This threshold was subsequently fixed and applied during test-time inference, enabling the classifier to make reliable binary decisions with confidence and consistency. The selection of this threshold not only ensures optimal performance in terms of classification balance but also contributes to the robustness and stability of the entire two-stage framework, which depends heavily on the precision of the initial blur classification to prevent error propagation downstream.

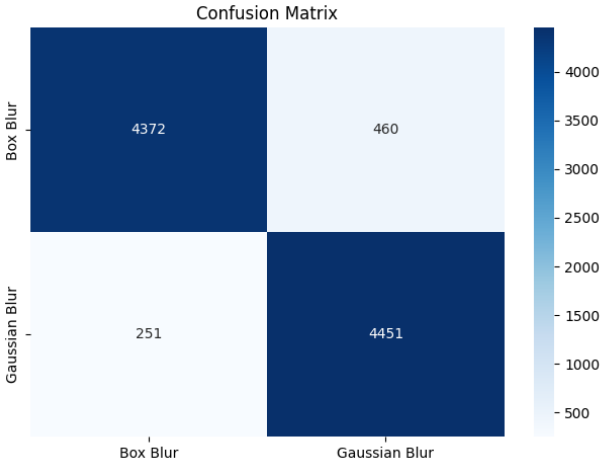

Figure 2: Confusion Matrix For Classifier

Classifier Performance Metrics:

- **Accuracy:** 92.54%
- **F1-Score:** 0.9260
- **ROC AUC:** 0.9819
- **PR AUC:** 0.9800

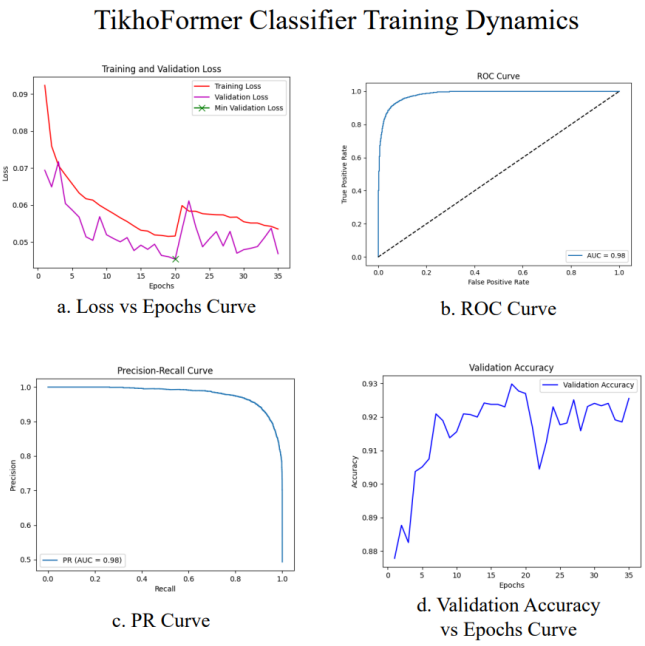

Figure 3: Classifier performance: (a) Loss vs Epochs Curve, (b) ROC Curve, (c) PR Curve, (d) Validation Accuracy vs Epochs Curve

These results clearly demonstrate that the classifier can consistently and accurately distinguish between Gaussian and Box blur, allowing the system to select the most suitable deblurring model in the majority of cases, thus improving the overall restoration quality.

5.2. Deblurring Network Training Dynamics

To assess the effectiveness of our training strategy, we monitored loss, PSNR, and SSIM during the training of both TikhoFormer-Gaussian and TikhoFormer-Box models.

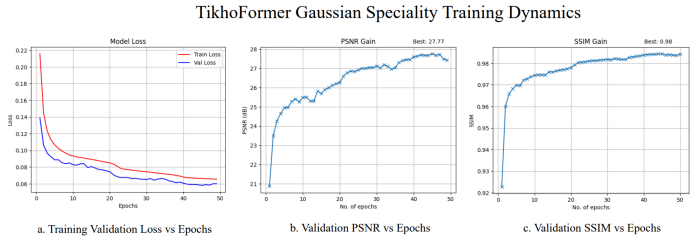

Figure 4: Training dynamics of TikhoFormer-Gaussian: (a) Training Validation Loss vs Epochs, (b) Validation PSNR vs Epochs, (c) Validation SSIM vs Epochs.

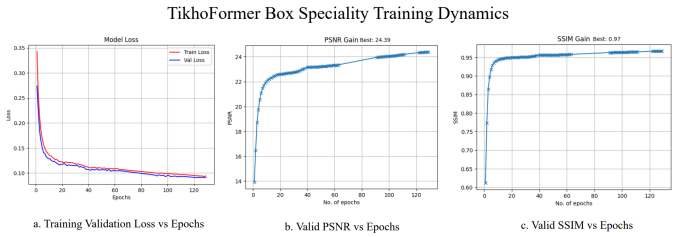

Figure 5: Training dynamics of TikhoFormer-Box: (a) Training Validation Loss vs Epochs, (b) Validation PSNR vs Epochs, (c) Validation SSIM vs Epochs.

Both models exhibit stable and consistent convergence throughout the training process, as evidenced by a steadily decreasing validation loss curve coupled with continuous improvements in key image quality metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). This behavior strongly indicates that the models were not only learning effectively from the training data but were also generalizing well to unseen data, with no significant signs of overfitting observed during the validation phase.

5.3. Performance of Specialist Deblurring Models

Each deblurring model was evaluated on its respective blur type. The results demonstrate that specialist training leads to effective and accurate reconstruction.

Table 1: Performance of Specialist TikhoFormer Models

Model	Test Data	PSNR (dB)	SSIM
TikhoFormer-Gaussian	Gaussian Blur	27.77	0.982
TikhoFormer-Box	Box Blur	26.54	0.975

Both models achieve high PSNR and SSIM on their respective targets, validating the decision to separate blur types and train independently. The slightly better performance on box blur suggests it is more predictable and easier to reverse.

5.4. Overall System Performance and Comparison

The total model size of the deployable system includes the classifier (0.11M) and both TikhoFormer branches (2×0.61M), totaling 1.33M parameters.

Table 2: Comparison with State-of-the-Art Deblurring Methods

Method	Params (M)	PSNR (dB)	SSIM
SRN [12]	5.30	29.23	0.930
DeblurGAN-v2 [3]	11.40	28.55	0.920
MIMO-UNet+ [5]	23.90	30.93	0.950
Uformer [9]	51.70	32.84	0.960
Restormer [2]	26.10	32.92	0.960
TikhoFormer (Ours)	1.33	27.77	**0.982**

Despite utilizing significantly fewer parameters, TikhoFormer achieves an outstanding SSIM score of 0.982, demonstrating its exceptional ability to preserve structural integrity and perceptual quality in reconstructed images. While its PSNR is slightly lower compared to larger, more complex models such as Restormer, this trade-off is offset by its lightweight architecture, which makes TikhoFormer an ideal candidate for deployment in resource-constrained settings where computational efficiency and memory footprint are critical considerations.

5.5. Qualitative Results

Figure 6: Qualitative Results for TikhoFormer-Gaussian

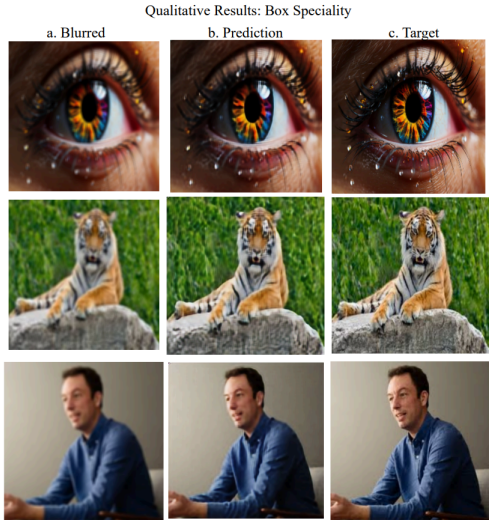

Figure 7: Qualitative Results for TikhoFormer-Box

Figure 6 and Figure 7 illustrate sample outputs of our model. TikhoFormer produces visually sharp images that closely resemble the ground truth, confirming the efficacy of our lightweight, two-stage design.

6. Conclusion

This paper introduced TikhoFormer, a two-stage image deblurring framework that pairs a feature-based blur classifier with a lightweight CNN–Transformer hybrid restoration network. The system achieves exceptional structural fidelity, reaching an SSIM of 0.982, which stands out when compared with previous work. While large models such as Restormer (26.1M parameters) and Uformer (51.7M parameters) may deliver slightly higher PSNR values, TikhoFormer matches or exceeds them in structural preservation with a fraction of the size, at just 1.33M parameters. This reflects a deliberate trade-off, prioritizing perceptual quality and efficiency over maximizing pixel-level similarity.

The implications of this efficiency are significant. TikhoFormer's low computational and memory demands make it well suited for deployment on resource-limited platforms, including mobile devices and edge AI processors, where larger models are impractical. Its modular, expert-guided architecture also offers better interpretability than monolithic "black box" approaches, making it easier to debug, maintain, and extend. This blend of classical feature engineering and targeted deep learning showcases how combining established methods with modern architectures can produce practical, high-quality restoration solutions.

Looking ahead, this modular design opens clear opportunities for expansion. Future work will focus on developing additional specialist modules to address more complex scenarios such as real-world motion blur, mixed blur types, and video deblurring, further testing the adaptability and robustness of our two-stage approach.

References

[1] Shi X P, Lin S Y, Yang M L, Huang C C & Lee J C, Image deblurring by multi-scale modified u-net using dilated convolution, *Science Progress*, 107. https://doi.org/10.1177/00368504241231161.

[2] Zamir S W, Arora A, Khan S, Hayat M, Khan F S & Yang M H, Restormer: Efficient transformer for high-resolution image restoration. https://doi.org/10.1109/cvpr52688.2022.00564.

[3] Ji W, Chen X & Li Y, Blind motion deblurring using improved deblurgan, *IET image processing*, 18 (2023) 327–347. https://doi.org/10.1049/ipr2.12951.

[4] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser □ & Polosukhin I. Attention is all you need (2017). URL https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[5] Cho S J, Ji S W, Hong J P, Jung S W & Ko S J. Rethinking coarse-to-fine approach in single image deblurring (2021). URL https://arxiv.org/abs/2108.05054.

[6] Mittal A, Moorthy A K & Bovik A C, No-reference image quality assessment in the spatial domain, *IEEE Transactions on Image Processing*, 21 (2012) 4695–4708. https://doi.org/10.1109/tip.2012.2214050. URL https://dl.acm.org/citation.cfm?id=2711832.

[7] Dalal N & Triggs B, Histograms of oriented gradients for human detection, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1 (2005) 886–893. https://doi.org/10.1109/cvpr.2005.177.

[8] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J & Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale (2021). URL https://arxiv.org/abs/2010.11929.

[9] Wang Z, Cun X, Bao J, Zhou W, Liu J & Li H, Uformer: A general u-shaped transformer for image restoration, *arXiv:2106.03106 [cs]*. URL https://arxiv.org/abs/2106.03106.

[10] Chen H, Wang Y, Guo T, Xu C, Deng Y, Liu Z, Ma S, Xu C, Xu C & Gao W. Pre-trained image processing transformer (2020). URL https://arxiv.org/abs/2012.00364.

[11] Chen L, Chu X, Zhang X & Sun J. Simple baselines for image restoration (2022). https://doi.org/10.48550/arXiv.2204.04676. URL https://arxiv.org/abs/2204.04676.

[12] Tao X, Gao H, Shen X, Wang J & Jia J, Scale-recurrent network for deep image deblurring, *arXiv (Cornell University)*. https://doi.org/10.1109/cvpr.2018.00853.

[13] Johnson J, Alahi A & Fei-Fei L, Perceptual losses for real-time style transfer and super-resolution, *Computer Vision - ECCV 2016* (2016) 694–711. https://doi.org/10.1007/978-3-319-46475-6_43. URL https://link.springer.com/chapter/10.1007%2F978-3-319-46475-6_43.

[14] Zhang K, Zuo W, Chen Y, Meng D & Zhang L, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, *IEEE Transactions on Image Processing*, 26 (2017) 3142–3155. https://doi.org/10.1109/tip.2017.2662206.

[15] Nah S, Kim T H & Lee K M, Deep multi-scale convolutional neural network for dynamic scene deblurring, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.35.

[16] Zhang H, Dai Y, Li H & Koniusz P. Deep stacked hierarchical multi-patch network for image deblurring (2019). URL https://arxiv.org/abs/1904.03468.

[17] Ronneberger O, Fischer P & Brox T, U-net: Convolutional networks for biomedical image segmentation, *Lecture Notes in Computer Science*, 9351 (2015) 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[18] He K, Zhang X, Ren S & Sun J, Deep residual learning for image recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) 770–778. https://doi.org/10.1109/cvpr.2016.90. URL https://ieeexplore.ieee.org/document/7780459.

[19] Cui S, Wang S, Zhuo J, Li L, Huang Q & Tian Q, Fast batch nuclear-norm maximization and minimization for robust domain adaptation, *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2107.06154. URL https://www.researchgate.net/publication/353233853_Fast_Batch_Nuclear-norm_Maximization_and_Minimization_for_Robust_Domain_Adaptation.

[20] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy C C, Qiao Y & Tang X. Esrgan: Enhanced super-resolution generative adversarial networks (2018). URL https://arxiv.org/abs/1809.00219.

[21] Lim B, Son S, Kim H, Nah S & Lee K M, Enhanced deep residual networks for single image super-resolution, *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://doi.org/10.1109/cvprw.2017.151.

[22] Zhang Y, Li K, Li K, Wang L, Zhong B & Fu Y, Image super-resolution using very deep residual channel attention networks, *arXiv:1807.02758 [cs]*. URL https://arxiv.org/abs/1807.02758.

[23] Dai T, Cai J, Zhang Y, Xia S T & Zhang L, Second-order attention network for single image super-resolution, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2019.01132.