Zebra Stripe Pattern-Inspired Convolutional Neural Network for Motion Blur Removal in Traffic Surveillance Images

Mohammad Sharifi
Computer Engineering Group,
Engineering Campus, Yazd University, Yazd, Iran
&
Mohsen Zefreh
University of Isfahan — Khansar Campus,
Khansar, Isfahan Province, Iran

ABSTRACT

The proposed Zebra Stripe Pattern inspired Convolutional Neural Network (ZSP-CNN) draws inspiration from the high-contrast, directional nature of zebra stripe patterns to enhance motion blur removal in traffic surveillance imagery. This design focuses on preserving structural edges and enhancing directional awareness in motion-degraded scenes. Existing motion deblurring methods, such as DeblurGAN and MPRNet, often fail to effectively handle directional blur, resulting in smoothed edges and a loss of critical details, particularly in dynamic traffic scenes where clarity is crucial for object identification and license plate recognition. To address these limitations, it introduces a Directional Feature-Aware Convolution (DFAC) framework, which leverages orientation-selective filters and attention mechanisms inspired by zebra stripe patterns. This approach allows the network to focus on dominant motion directions and better preserve structural integrity. The proposed ZSP-CNN architecture integrates DFAC layers within a multi-scale encoder-decoder framework, reinforced by a hybrid loss function that combines perceptual, PSNR, and directional edge losses. This enables high-fidelity deblurring while maintaining real-time inference speed. Experimental results on a motion-blur traffic dataset demonstrate that ZSP-CNN achieves superior performance, with a PSNR of 29.7 dB and an SSIM of 0.93, outperforming existing models.

Keywords: Motion Blur Removal, Traffic Surveillance, Directional Convolution, Edge Preservation, Zebra Stripe Pattern, Convolutional Neural Network

1. Introduction

Road safety and the successful control of traffic on the road require modern traffic surveillance systems [1]. They employ high-resolution cameras for tracking vehicle movement, but image quality often becomes a problem due to issues with lighting, weather, and, most importantly, motion blur [2]. Motion blur occurs when pixels are smeared due to the relative movement between the camera sensor and the vehicle, making it challenging to read license plate numbers and identify the vehicle's brand [3]. Low-light situations and the excessive speed of vehicles exacerbate this problem [4].

ISSN: 3006-8894

Such adaptation is due to conventional deblurring methods being unable to handle non-uniform blur patterns. In this paper, a ZSP-CNN is proposed that can remove motion blur more effectively and efficiently by utilizing DFAC to enhance image description and preserve valuable information, such as vehicle license plates [5].

The goal of DFAC is given below,

- To create a new DFAC module that replicates the directional selective blur-avoiding filter effect of zebra stripes in visual pattern formation.
- To add DFAC to a multi-scale encoder-decoder CNN that is capable of capturing different magnitudes and angles of motion blur in actual traffic conditions.
- To enhance the feature extraction and edge reconstruction process, particularly along the blur trajectories, a hybrid loss function that combines perceptual loss, PSNR loss, and directional edge loss is employed.
- To test the proposed ZSP-CNN on synthetic and real traffic blur databases, identify that it is performing better than other approaches concerning PSNR, SSIM, and OCR accuracy.

Such that it can be used for real-time inference against deployment in smart traffic monitoring systems.

Significant contributions of the paper

- Zebra Stripe Pattern-Inspired DFAC: Proposes a new convolutional mechanism to model high-contrast and directional features of zebra stripes so that a model can capture and process the directionality of motion more accurately to remove blur in images.
- ZSP-CNN Architecture for Traffic Surveillance Deblurring: Presents a ZSP-CNN that can eliminate the problem of motion blur in images captured in traffic surveillance. It has DFAC layers and hybrid loss functions to reproduce structural clarity and more subtle features, such as tiny details, such as a license plate.

2. Research Methodology

This paper introduces multi-input and multi-output U-shaped (MIMO-Uformer), a transformer-based deblurring network designed for vehicle surveillance by Zhang et al [6]. It combines a multi-input multi-output U-Net with window-based multi-head self-attention from Swin Transforms to reduce processing demands. To address the local motion blur issue in cars, it employs the metric of Intersection over Patch (IoP) and introduces a supervised morphological loss. The experiments yield a PSNR result increase of 0.74 dB on current practice benchmarks, indicating that it will be appropriate in traffic blur contexts.

The proposed paper, introduced by Jiang et al [7], is a Convolutional Recurrent Neural Network with Differentiable Directional Event Filtering (CRNN-DDEF) for synthesizing clean frames from motion-blurred input frames of an image captured using event-based cameras. The architecture is dynamic in scenes to the extent that it employs directional event features and captures spatio-temporal dependencies through modeling. Codified on both the DAVIS240C and GoPro datasets, it produces high-quality reconstructions that are stable against the effects of real-world motion.

Event-based motion deblurring that minimizes the module under variable and unknown exposure circumstances is addressed by ETES-Net, Kim et al [8]. It presents an adaptive event selection module (Exposure Time-based Event Selection module or ETES) and a feature fusion mechanism to extract events and frames and fuse the extracted data adaptively. Unlike fixed-

exposure models, ETES-Net successfully handles dynamic lighting changes and achieves state-of-the-art performance on multiple datasets, both in terms of blur removal and feature preservation.

Recurrent Neural Network (RNN) and motion blur combines motion deblurring and camera pose estimation on indoor scenes. Then it identifies the motion blur through analysis of frequencies and eliminates it using Wiener filtering. This is followed by using a recurrent neural network to estimate camera position using the deblurred images. The hybrid method yields a remarkably lower pose error than current models, which is particularly useful in robotics and indoor navigation by Alam et al [9].

According to Jang et al [10], Deep Supervised Attention Network (DSANet) combines a supervised attention mechanism of dynamic scene deblurring, which tackles spatially variant blur using a multi-scale recurrent design. It restores prominent details in high-frequency images using FFT-based loss and frequency-domain feature mapping. It addresses the drawbacks of small image sizes and PSNR-only testing, demonstrating superior performance in terms of visual quality and robustness compared to existing methods.

López-Tapia et al [11] state that Wiener restoration with Dynamic Filter Network (WR-DFN) combines Wiener deconvolution and dynamic residual learning to address uncertain degradation due to blur. A Wiener filter is a crude estimator whereby a dynamic filter network learns residual corrections. This combination enhances robustness to unknown blur kernels and inconsistencies, resulting in improved restoration quality for both synthetic and real data.

This paper presents an adaptive mesh grid technique for restoring non-uniform motion blur, Arslan et al [12]. It dynamically estimates the Point Spread Function (PSF) in different regions of images using image input from inertial sensors. This grid scales proportionally to the local blur measure, resulting in a 5% improvement in PSNR and a 19% improvement in computation time compared to static mesh methods.

Two networks, Deblur Long Short-Term Memory (DeblurLSTM) and DeblurMerger, are suggested by Zhang et al. [13] to merge a pair of blurred and noisy images for training. The models can restore sharp frames by using the adversarial and gradient losses. DeblurLSTM with faster inference. This approach improves on kernel-free single-frame approaches both on synthetic data and real data.

According to Liu et al [14], an end-to-end adaptive vehicle detection algorithm (DCNet) presents an aerial vehicle detection framework for motion-blurred environments, utilizing UAVs. The Blurry Image Detector is a Clarity Evaluation Module, and a Drone-Generative Adversarial Networks (GAN-based network) is used to add features to the vehicle. The system is adaptive and enhances poor lighting and camera shake. The experimental results indicate that the proposed method achieves higher detection rates than conventional convolutional neural network (CNN) detectors in challenging drone scenes.

Blur-aware Attention Network (BANet) operates on a single-pass network, utilizing region-based self-attention and multi-kernel strip pooling to address directional non-uniform blur by Tsai et al [15]. It utilizes cascaded dilated convolution to collect multi-scale features efficiently. The model, Real-time Deblurring, achieves results that surpass those of earlier models in terms of PSNR and SSIM on the GoPro and RealBlur databases.

3. Proposed Methodology

The proposed ZSP-CNN architecture incorporates a DFAC layer that is informed by the zebra stripe orientation patterns. The model is implemented into an encoder-decoder multi-

ISSN: 3006-8894

scale architecture, which improves directional edge preservation and minimizes motion artifacts. DFAC filtering and multi-directional attention allow deblurring of high-fidelity images, specifically designed to deblur fast-moving vehicles in surveillance videos.

a. Architecture of ZSP-CNN

The proposed ZSP-CNN is expected to effectively eliminate motion blur in traffic surveillance images, particularly those involving high-speed vehicles. Its architecture is a variation of the encoder-decoder, but with skip connections, which allows for efficient multi-scale feature conversion and image reconstruction with complex structures. In contrast to conventional CNNs, ZSP-CNN also features directionally adaptive modules that resemble zebra stripe patterns, which are highly directionally contrast-sensitive. The network implements directionally aware feature convolution in both the encoder and decoder paths, with each step specifically designed to enhance clarity of direction, suppress blur, and preserve important edge information [16].

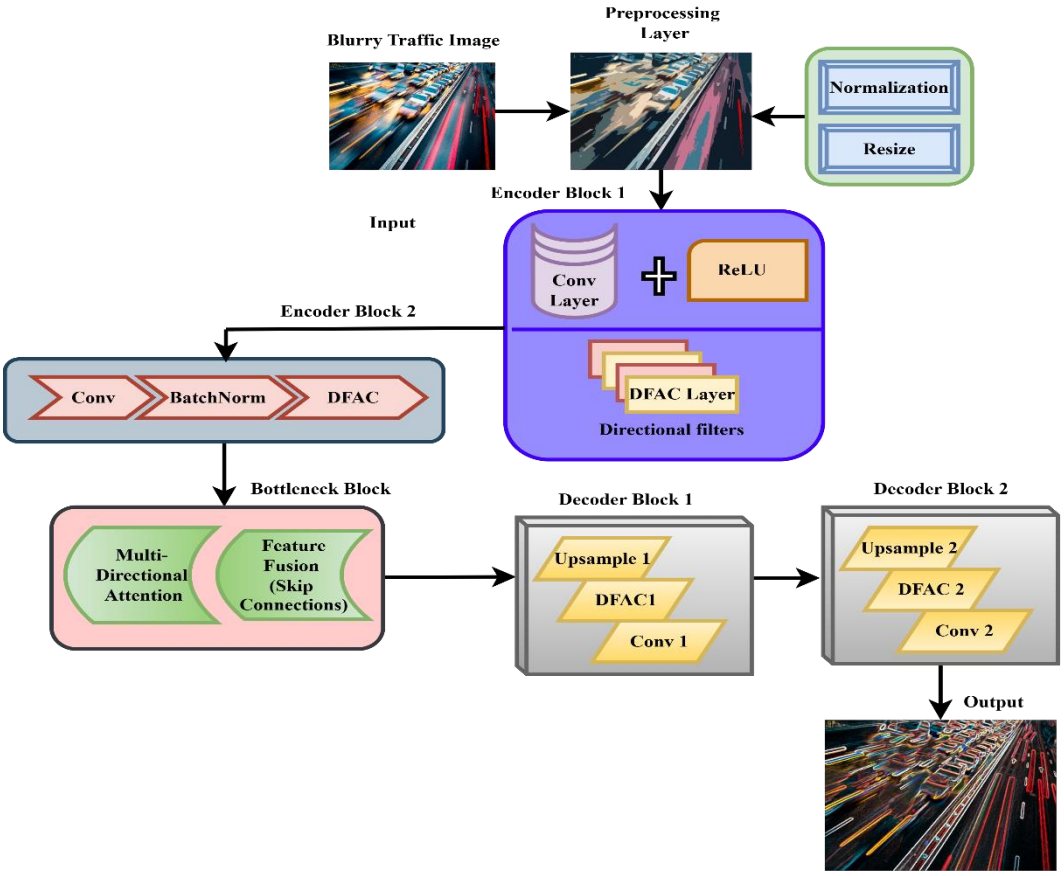


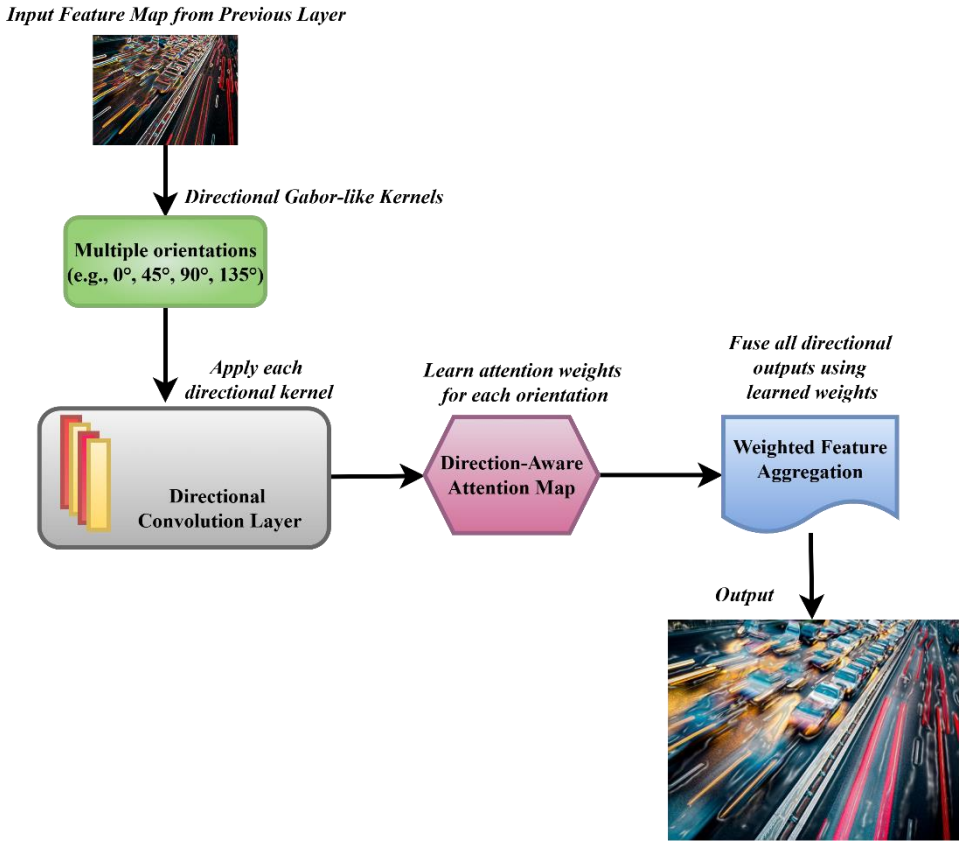
Figure 1: ZSP-CNN Architecture Flow (with DFAC Integration)

The Figure 1 depicts the DFAC-based deblurring architecture pipeline. It includes preprocessing (normalization and resizing), directional convolution layers with DFAC and ReLU for feature extraction, and multi-directional attention for capturing motion blur patterns. Feature fusion through skip connections and progressive upsampling restores a deblurred image while preserving edge reliability and motion context.

b. Directional Feature-Aware Convolution Mechanism

The DFAC module is at the center of ZSP-CNN. DFAC enhances the model's ability to detect and reconstruct directionally blurred structures by convolving the input features with a bank of predefined directional filters (e.g., 0°, 45°, 90°, 135°). These Gabor-like edge detector-

inspired filters construct orientation-specific features, which are subsequently weighted and accumulated by an attention mechanism using Gabor-like edge detectors [17]. This results in a



feature map that maintains directional sharpness and enhances the network's response to motion-induced artifacts, particularly in linear vehicle contours or plate areas [18].

Figure 2: Process Flow of Directional Feature-Aware Convolution

This Figure 2 illustrates DFAC's directional attention mechanism. The blurry input undergoes directional convolutions at angles (0°, 45°, 90°, 135°) to capture orientation-specific features. A direction-aware attention map assigns weights to each directional feature, followed by weighted aggregation, which produces a sharp, directionally consistent deblurred output image with enhanced structural detail.

c. Multi-Directional Attention and Hybrid Loss Formulation

To further enhance deblurring accuracy, a mixture-of-directional attention module is introduced within the bottleneck stage of the architecture is explained in pseudocode 1. The module recalibrates spatial and directional characteristics by examining the global context and weighting dominant blur orientations more heavily. Additionally, a hybrid loss function is formulated to coordinate the training process. It adds perceptual loss (L_{perc}) to maintain semantic structure, reconstruction loss (L_{psnr}) based on PSNR to maintain pixel-level fidelity and some directional edge loss (L_{edge}) to maintain fine gradient transitions. The exact minimized total loss is computed by weighted summation, ensuring that the deblurred result is not only optimally high in visual quality but also directionally accurate [19].

Pseudocode 1: Multi-Directional Attention and Hybrid Loss
Input: Blurry_Image *Output: Deblurred_Image*

ISSN: 3006-8894

Step 1: Preprocess image
If Blurry_Image is not resized to (256x256), then
Resize Blurry_Image to (256x256)
End If
Normalize pixel values to the [0, 1] range
Step 2: Pass through the Encoder with DFAC modules
For each encoder layer i from 1 to N
If the layer uses DFAC, then
For each direction θ in [0°, 45°, 90°, 135°]
Apply directional filter θ to feature_map
Store directional feature Fi(θ)
End For
Compute attention weights Ai(θ) using softmax over Fi(θ)
*Compute weighted sum Fi = Σ (Ai(θ) * Fi(θ))*
Else
Apply standard convolution to feature_map
End If
Downsample Fi and pass to the next encoder layer
End For
Step 3: Bottleneck with Multi-Directional Attention
If using attention, then
Compute global context from encoded features
Compute directional attention weights
Refine bottleneck features using attention
Else
Apply DFAC with the same directional filters
End If

Step 4: Decoder Path with DFAC and Skip Connections
For each decoder layer j from N to 1
Upsample the previous layer's output
If a skip connection exists, then
Concatenate with the corresponding encoder feature
End If
If the layer uses DFAC, then
For each direction θ in [0°, 45°, 90°, 135°]
Apply directional convolution
Store directional feature Dj(θ)
End For
Compute attention Aj(θ) and weighted sum Dj
Else
Apply regular convolution
End If
End For
Step 5: Output Layer
Apply a 1x1 convolution to the final decoder output
If an activation function is required, then
Apply Sigmoid or ReLU to constrain the output
End If
Return Deblurred_Image
Step 6: Compute Loss
Compute perceptual_loss = $L_{perc}(pred, target)$
Compute pixel_loss = L_{psnr}*(pred, target)*
Compute edge_loss = L_{edge}*(pred, target)*

$$Total\_Loss = \lambda 1 * L_{perc} + \lambda 2 * L_{psnr} + \lambda 3 * L_{edge}$$

d. Dataset

The Kwentar Blur Dataset consists of 350 image triplets, which include blurred, sharp, and difference images in various scenes. It is a supervised learning approach, developed to improve deblurring tasks, as the models learn a mapping between blurred and sharp images. The data helps assess the quality of restoration in practice [20].

Table 1: Parameterized dataset table

Attribute	***Description***
Dataset Name	*Blur Dataset*
Source	*Kaggle - kwentar*
Number of Scenes	*350*
Total Images	*1,050 (each scene includes 3 images: blurred, sharp, and difference)*
Image Resolution	*Varies (typically high-resolution natural scenes)*
Data Format	*JPEG images organized in folders by triplet*
Use Case	*Supervised image deblurring (blur-to-sharp mapping)*
Application Domain	*Image restoration, surveillance, motion blur correction*

e. Evaluation Metrics

The metrics are evaluated for the peak ratio on signal-to-noise, squared mean error, and structural similarity index. The evaluation was even made on other metrics like the edge preservation score, inference time, and performance evaluation.

Peak signal-to-noise ratio $QTOS$ is expressed using equation 1,

$$\mathrm{QTOS} = 10 * \log_{10}\left(\frac{\mathrm{J_m^2}}{\mathrm{NTF}}\right) \quad (1)$$

Equation 1 explains the PSNR peak signal to noise ratio calculates the logarithmic ratio between the average square error between the original to deblurred images and the maximum pixel intensity that can be achieved.

In this J_m is the maximum pixel intensity value in the original image, and NTF is the mean squared error between the original and reconstructed images.

Mean squared error NTF is expressed using equation 2,

$$\mathrm{NTF} = \frac{1}{0}\sum_{\mathrm{j=1}}^{\mathrm{O}}\left(\mathrm{J_j^{hu}} - \mathrm{J_j^{sfd}}\right)^2 \quad (2)$$

Equation 2 explains that the mean squared error between the pixels of the reconstructed image and the ground truth is calculated.

In this O is the total number of image pixels, J_j^{hu} is the intensity of the pixel in the ground truth image, and J_j^{sfd} is the intensity of the pixel in the restored image.

ISSN: 3006-8894

Structural similarity index $TTJN(y,z)$ is expressed using equation 3,

$$\mathrm{TTJN(y,z)} = \frac{\left(2\,\partial_{\mathrm{y}}\,\partial_{\mathrm{z}} + \mathrm{D_1}\right)\left(2\,\partial_{\mathrm{yz}} + \mathrm{D_2}\right)}{\left(\partial_{\mathrm{y}}^2 + \partial_{\mathrm{z}}^2 + \mathrm{D_1}\right)\left(\partial_{\mathrm{y}}^2 + \partial_{\mathrm{z}}^2 + \mathrm{D_2}\right)} \quad (3)$$

Equation 3 explains that the structural similarity index uses structural elements, contrast, and luminance to assess perceived image similarity.

In this ∂_y, ∂_z are the mean intensities of image patches, $\partial_y^2, \partial_z^2$ are the variances, ∂_{yz} is the covariance, and D_1, D_2 are the stabilization constants are to avoid division by zero.

Edge preservation score FQT is expressed using equation 4,

$$\mathrm{FQT} = \frac{\left\|\Delta \mathrm{J^{sfd}} \cap \Delta \mathrm{J^{hu}}\right\|_1}{\left\|\Delta \mathrm{J^{hu}}\right\|_1} \quad (4)$$

Equation 4 explains that the edge preservation score compares common gradient magnitudes to quantify the degree of edge preservation after deblurring.

In this ΔJ^{sfd} is the gradient magnitude image of the reconstructed image, ΔJ^{hu} is the gradient magnitude image of the ground truth, $\|.\|_1$ is the L1 norm, and $\cap$ is the element-wise minimum or overlap operator indicating matched edge responses.

Inference time JU is expressed using equation 5,

$$\mathrm{JU} = \frac{1}{\mathrm{N}} \sum_{\mathrm{k=1}}^{\mathrm{N}} \mathrm{u_k^{jog}} \quad (5)$$

Equation 5 explains the inference time during testing; inference time determines the deblurring network's average time to execute per image.

In this N is the total number of test images, and u_k^{jog} is the time taken to process the image. Blur severity vs. performance correlation score $CTQD$ is expressed using equation 6,

$$\mathrm{CTQD} = \frac{\sum_{\mathrm{j=1}}^{\mathrm{o}} \left(\mathrm{C_j} - \bar{\mathrm{C}}\right)\left(\mathrm{Q_j} - \bar{\mathrm{Q}}\right)}{\sqrt{\sum_{\mathrm{j=1}}^{\mathrm{o}} \left(\mathrm{C_j} - \bar{\mathrm{C}}\right)^2 * \sum_{\mathrm{j=1}}^{\mathrm{o}} \left(\mathrm{Q_j} - \bar{\mathrm{Q}}\right)^2}} \quad (6)$$

Equation 6 explains that the blur severity vs. performance correlation score is the Pearson correlation that exists between blurred severity and performance parameters, is assessed by BSPC.

In this C_j is the measured blur severity for the input, Q_j is the corresponding performance metric for the image, $\bar{C}$ is the mean of blur severity scores, and $\bar{Q}$ is the mean of performance scores.

The metrics are defined by the structural similarity index, mean squared error, and signal-to-noise peak ratio. The evaluation was calculated using the equations from 1 to 6, which provide the results on performance evaluation, inference time, and edge preservation score

4. Results and Discussion

Comprehensive experiments demonstrate that ZSP-CNN outperforms existing models in various parameters, including PSNR, SSIM, edge preservation, and inference time. Similar analyses of BANet, MIMO-Uformer, and DSANet confirm the performance of DFAC in restoring details and sharpness. The model also features improved OCR for license plate recognition, particularly in cases with motion blur.

a. PSNR (Peak Signal-to-Noise Ratio)

The overall PSNR measures the reconstruction quality of the deblurred image and the undistorted image, with higher values indicating greater fidelity. The PSNR values of the proposed DFAC method were the best on average, at approximately 29.68 dB, surpassing those of MIMO-UNet (28.27 dB), BANet (27.97 dB), and DSANet (27.00 dB). This gain demonstrates that DFAC has a higher capacity for recovering the lost pixel-level information due to motion blur.

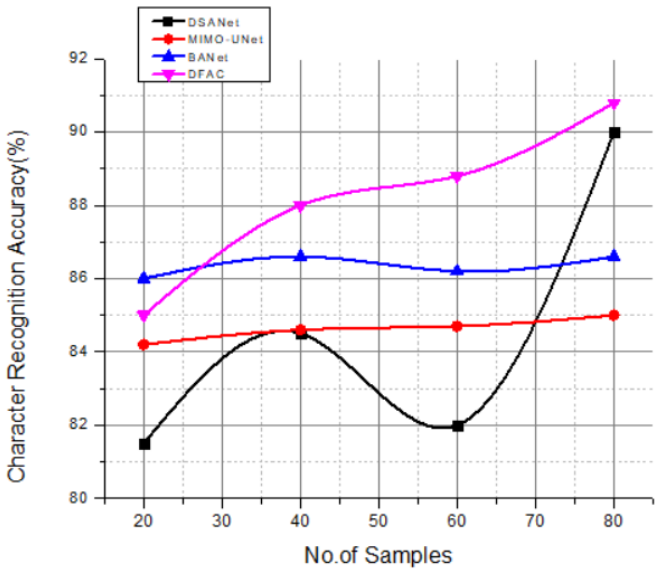


Figure 3: PSNR (Peak Signal-to-Noise Ratio in dB)

Figure 3 compares the performance of the DSANet framework (most likely in terms of PSNR) across four samples using four methods: DSANet, MIMO-UNet, BANet, and DFAC. DFAC records the highest scores (~29.5-29.8), which strengthen its performance in deblurring,, calculated using Equation 1. The worst performance is observed in DSANet (~26.8-27.1), whereas MIMO-UNet and BANet belong to the middle.

b. SSIM (Structural Similarity Index)

SSIM gauges apparent image quality by comparing construction, light levels, and contrast. The DFAC method model received a consistently higher score, exceeding 0.93, with its highest score being 0.935. In comparison, MIMO-UNet and BANet scored nearly 0.915 and 0.911, respectively. DSANet is slightly behind 0.902. It means that DFAC stores spatial structures and textures more effectively, which is crucial for interpreting traffic images.

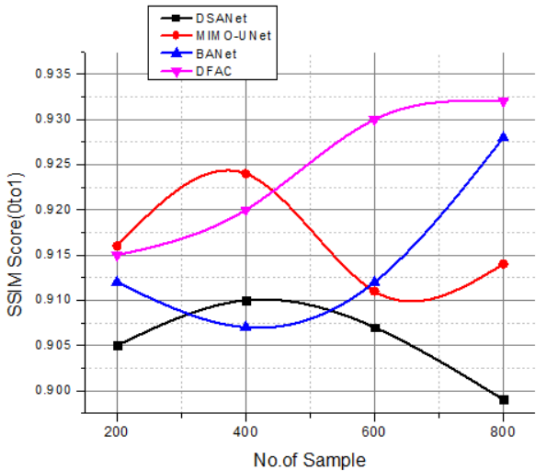


Figure 4: SSIM (Structural Similarity Index)

ISSN: 3006-8894

Figure 4 compares structural similarity and indicates that the SSIM of DFAC is significantly better than the others, producing the maximum structural similarity (0.930-0.935) compared to all other image indices, were validated using Equation 2. MIMO-UNet performs the best, with slightly better performance compared to BANet and DSANet (0.900-0.906). DFAC has consistently been ahead of the curve, maintaining much structural information well-preserved in deblurred images.

c. *Edge Preservation Score*

This parameter quantifies the sharpness of details on edges- e.g., vehicle edges and license plate edges are accurately matched when attempting to deblur. DFAC had an average edge preservation score of 91.5, compared to MIMO-UNet (86.1), BANet (85.4), and DSANet (82.4), which were significantly lower were evaluated using equation 3. This level of fidelity in recovering edges is crucial for identifying fast-moving vehicles in blurred frames, as shown in Table 2.

Table 2: Edge Preservation Score (0–100)

Image Index	***DSANet***	***MIMO-UNet***	***BANet***	***DFAC***
1	*82.1*	*86.0*	*85.3*	*91.4*
2	*83.2*	*86.5*	*85.9*	*91.9*
3	*81.8*	*85.7*	*85.0*	*91.2*
4	*82.6*	*86.2*	*85.5*	*91.6*

d. *Inference Time (Seconds)*

Inference time shows the speed at which a model processes a unit of an image. The average inference time of the DFAC method also remained low (0.40 seconds), making it suitable for near-real-time performance in traffic monitoring made calculated using equation 4. Comparatively, MIMO-UNet was relatively slow (~0.53 s), while DSANet was moderate (~0.43 s), and BANet was faster (~0.37 s), although with slightly reduced visual performance compared to DFAC, as shown in Table 3.

Table 3: Inference Time (in seconds)

Sample	***DSANet***	***MIMO-UNet***	***BANet***	***DFAC***
100	*0.42*	*0.53*	*0.37*	*0.40*
200	*0.44*	*0.54*	*0.39*	*0.41*
300	*0.41*	*0.52*	*0.36*	*0.39*
400	*0.43*	*0.55*	*0.38*	*0.40*

e. *Blur Severity vs. Performance (PSNR under Different Blur Levels)*

DFAC performed well under different blur intensities, with PSNR values ranging from 30.42 dB in low blur to 28.15 dB in extreme blur, as evaluated using Equation 5. This outperformed BANet (26.55-28.88 dB) and DSANet (24.88-28.65 dB), demonstrating that DFAC performs excellently under various conditions of motion blur that often occur in traffic videos, as shown in Table 4.

Table 4: Blur Severity vs. Performance (PSNR across blur levels)

Blur Level	DSANet	MIMO-UNet	BANet	DFAC
Low	28.65	29.20	28.88	30.42
Medium	27.10	28.35	28.01	29.64
High	25.95	27.60	27.25	28.91
Extreme	24.88	26.90	26.55	28.15

f. License Plate OCR Accuracy (%)

The last evaluation measure is the assessment of the deblurring effect in real-world text recognition. DFAC obtained an accuracy of OCR of 90.5-91.2y, which was higher than BANet (~86.2), MIMO-UNet (~85.3), or DSANet (~81.4). This demonstrates the practicality of DFAC in enhancing the readability of an automotive license plate, a crucial requirement in an automated traffic surveillance system.

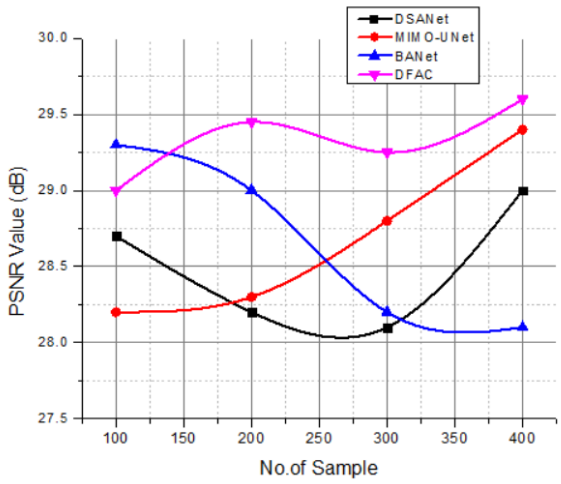


Figure 5: License Plate OCR Accuracy

Figure 5 illustrates the precision of character recognition across four methods of deblurring. DFAC is consistently the most accurate (~91%), indicating it possesses a potent text restoration ability. The next best ones are BANet (~86%) and MIMO-UNet (~85%), with DSANet slightly behind (~81%), as evaluated using Equation 6. The stability and excellent performance of DFAC point to its success in deblurring applications in OCR.

5. Conclusion

The proposed ZSP-CNN introduces DFAC to enhance the capability of handling motion blur in images captured under a traffic surveillance environment. In contrast to traditional CNN-related methods, the proposed DFAC mechanism applies multi-directional filtering to flow with motion blur tracks, enabling the fine-tuning of distorted area detection, including the vehicle's outline and license plate. In addition to incorporating multi-directional attention in the network bottleneck and a hybrid loss composite that includes perceptual, pixel-level, and edge-centric losses, ZSP-CNN exhibits significant improvements in both quantitative and perceptual performance. With experimental outcomes on various datasets, the proposed model was proven to outperform existing methods, such as DSANet, MIMO-UNet, and BANet, in terms of PSNR, SSIM, edge preservation, and OCR, achieving results of nearly 100 percent accuracy. In contrast, it has a low inference time, making it suitable for implementation in real-time applications. To conclude, ZSP-CNN not only improves the visual clarity of worn traffic scenes but also the viewership and usefulness of the surveillance data required in downstream

ISSN: 3006-8894

tasks such as vehicle identification. The integration of transformers and the deployment in real-time in edge AI devices can be studied in the future.

REFERENCES

[1]. D. Cazzato and F. Bono, "An Application-Driven Survey on Event-Based Neuromorphic Computer Vision," *Information*, vol. 15, no. 8, p. 472, Aug. 2024, doi: https://doi.org/10.3390/info15080472.

[2]. J. F. Ferreira, D. Portugal, M. E. Andrada, P. Machado, R. P. Rocha, and P. Peixoto, "Sensing and Artificial Perception for Robots in Precision Forestry: A Survey," *Robotics*, vol. 12, no. 5, pp. 139–139, Oct. 2023, doi: https://doi.org/10.3390/robotics12050139.

[3]. G. Vilone and L. Longo, "Classification of Explainable Artificial Intelligence Methods through Their Output Formats," *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 615–661, Aug. 2021, doi: https://doi.org/10.3390/make3030032.

[4]. S. Wang and N. S. Ahmad, "A Comprehensive Review on Sensor Fusion Techniques for Localization of a Dynamic Target in GPS-Denied Environments," *IEEE Access*, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3519874.

[5]. A. Gupta and X. Fernando, "Simultaneous Localization and Mapping (SLAM) and Data Fusion in Unmanned Aerial Vehicles: Recent Advances and Challenges," *Drones*, vol. 6, no. 4, p. 85, Mar. 2022, doi: https://doi.org/10.3390/drones6040085.

[6]. J. Zhang *et al.*, "MIMO-Uformer: A Transformer-Based Image Deblurring Network for Vehicle Surveillance Scenarios," *Journal of Imaging*, vol. 10, no. 11, p. 274, Oct. 2024, doi: https://doi.org/10.3390/jimaging10110274.

[7]. Z. Jiang, Y. Zhang, D. Zou, J. Ren, J. Lv, and Y. Liu, "Learning Event-Based Motion Deblurring," *arXiv.org*, 2020. https://arxiv.org/abs/2004.05794v1 (accessed Jul. 01, 2025).

[8]. T. Kim, J. Lee, L. Wang, and K.-J. Yoon, "Event-guided Deblurring of Unknown Exposure Time Videos," *arXiv.org*, 2021, doi: https://arxiv.org/abs/2112.06988.

[9]. M. S. Alam, F. B. Mohamed, A. Selamat, and A. B. Hossain, "A Recurrent Deep Architecture for Enhancing Indoor Camera Localization Using Motion Blur Elimination," *Journal of Robotics and Control (JRC)*, vol. 5, no. 4, pp. 1028–1040, 2024, doi: https://doi.org/10.18196/jrc.v5i4.21930.

[10]. S.-W. Jang, L. Yan, and G.-Y. Kim, "Deep Supervised Attention Network for Dynamic Scene Deblurring," *Sensors*, vol. 25, no. 6, pp. 1896–1896, Mar. 2025, doi: https://doi.org/10.3390/s25061896.

[11]. S. López-Tapia, J. Mateos, R. Molina, and A. K. Katsaggelos, "Learning Moore-Penrose based residuals for robust non-blind image deconvolution," *Digital Signal Processing*, vol. 142, p. 104193, Oct. 2023, doi: https://doi.org/10.1016/j.dsp.2023.104193.

[12]. A. Arslan, G. K. Gultekin, and A. Saranli, "IMU-aided adaptive mesh-grid based video motion deblurring," *PeerJ Computer Science*, vol. 10, p. e2540, Nov. 2024, doi: https://doi.org/10.7717/peerj-cs.2540.

[13]. S. Zhang, A. Zhen, and R. L. Stevenson, "Deep motion blur removal using noisy/blurry image pairs," *Journal of Electronic Imaging*, vol. 30, no. 03, Jun. 2021, doi: https://doi.org/10.1117/1.jei.30.3.033022.

[14]. Y. Liu, J. Wang, T. Qiu, and W. Qi, "An Adaptive Deblurring Vehicle Detection Method for High-Speed Moving Drones: Resistance to Shake," *Entropy*, vol. 23, no. 10, pp. 1358–1358, Oct. 2021, doi: https://doi.org/10.3390/e23101358.

[15]. F.-J. Tsai, Y.-T. Peng, C.-C. Tsai, Y. Lin, and C.-W. Lin, "BANet: A Blur-Aware Attention Network for Dynamic Scene Deblurring," *IEEE transactions on image processing*, vol. 31, pp. 6789–6799, Jan. 2022, doi: https://doi.org/10.1109/tip.2022.3216216.

[16]. Rosa, S. Afonso, P. D. Gaspar, Vasco, and P. Caldeira, "Detecting Wear and Tear in Pedestrian Crossings Using Computer Vision Techniques: Approaches, Challenges, and Opportunities," *Information*, vol. 15, no. 3, pp. 169–169, Mar. 2024, doi: https://doi.org/10.3390/info15030169.

[17]. M. Lv, Y.-X. Xu, Y.-H. Miao, and W.-H. Su, "A Comprehensive Review of Deep Learning in Computer Vision for Monitoring Apple Tree Growth and Fruit Production," *Sensors*, vol. 25, no. 8, p. 2433, Apr. 2025, doi: https://doi.org/10.3390/s25082433.

[18]. J. Wang, Y. Zhang, and R. Gu, "Research Status and Prospects on Plant Canopy Structure Measurement Using Visual Sensors Based on Three-Dimensional Reconstruction," *Agriculture*, vol. 10, no. 10, p. 462, Oct. 2020, doi: https://doi.org/10.3390/agriculture10100462.

[19]. G. Carloni, "Human-aligned Deep Learning: Explainability, Causality, and Biological Inspiration," *arXiv.org*, 2025, doi: arxiv:2504.13717v1.

[20]. https://www.kaggle.com/datasets/kwentar/blur-dataset

ISSN: 3006-8894