Landmark Guided HINet for Facial Image Deblurring

Rutuja Dautpure[1], Sakshi Powar[2], Sayali Khedkar, Vaishnavi Shelar, Amit Joshi, and Soma Ghosh

COEP Technological University
Pune, India
dautpurerv22.comp@coeptech.ac.in

Abstract. Facial images captured in real life often get blurred due to motion blur or poor focus. Facial image deblurring plays an important role in surveillance, authentication and biometric applications. Although conventional Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN) based deblurring methods achieve promising results, they fail to preserve facial geometry because normalization distorts important facial details. To resolve this limitation, a lightweight data-driven Landmark Guided Half Instance Normalization (HIN) architecture is proposed, which enhances the Half Instance Normalization Network (HINet) deblurring model by adding facial landmarks as structural guidance. In the proposed approach, facial landmarks act as guidance that inform the network which regions contain critical facial details. HIN is applied to non-facial areas, which then helps the model to preserve important features like eyes, nose and mouth. Also, landmark prediction branch is added with a heatmap-based landmark loss. Experimental results show that proposed model produces sharper and structurally consistent face restorations. Quantitative experiments show that the model achieves a Peak Signal-to-Noise Ratio (PSNR) of 26.2 and Structural Similarity Index Measure (SSIM) of 0.91, which indicates improvements. These results demonstrate that integrating facial priors improves recognition accuracy, supporting intelligent and sustainable surveillance systems.

Keywords: Normalisation, HINet, Facial Landmarks, CNN, Image Deblurring.

1 Introduction

Images serve as critical data sources in various fields such as medical, surveillance, remote sensing and authentication. In real world, intelligent surveillance systems are used for public safety and urban monitoring. Facial images in CCTV environment suffer degradation due to motion blur caused by movements and low light conditions. Because of this there is chance of losing significant landmarks of the image which leads to weakening tasks like face recognition and authentication. Images used for facial deblurring task contains details like eyes,

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_76

nose and lips that need to stay consistent and realistic. Facial image deblurring comes with challenges, as it requires preserving important facial details like eyes, nose, lips and facial contours. These facial landmarks are essential information for face recognition and authentication tasks.

Traditional approaches to the facial deblurring task have used semantic masks as network input to provide structural prior information about different facial parts [1] [2]. Even though these methods improved overall reconstruction quality but they often failed to satisfactorily restore fine details around facial landmarks. This limitation occurs because the pixel area of these critical facial parts is relatively small, and semantic masks lack texture-related details. To address this, recent methods have included landmark-related loss functions that guide the network to accurately detect and reconstruct facial landmark points.

Recent experiments have shown that CNN and GAN were found to have performed exceptionally well. In CNN, various features are extracted like edges and texture through convolution layers and are reconstructed again to remove the blur patterns. In the past few years, various deep learning architectures have been proposed for image deblurring. Tian et al. proposed an experimental deep learning-based model along with mathematical representation to generate clear images from deblurred images [3]. Zhang et al. developed DFPG-net, a GAN based model [4]. It consists of two parts, first part extracts the important facial features and other part uses these features to make images sharp. Along with this, SFT and SSFT is added to improve fine details. Wang et al. used an unsupervised GAN based model which uses a Scale Adaptive Attention Module (SAAM) which allows the model to focus on details at different scales [5]. So, Jung et al. explored a CNN based method like U-shaped Convolutional Neural Network (U-NET) and encoder-decoder models [6]. With this method there is no need to calculate the blur kernel and it directly gives the sharpened image. Ye et al. proposed SIUN, a scale-iterative upscaling network that progressively refines blurred images through multiple U-Net stages with residual dense blocks [7]. Lokesh worked on traditional deblurring method which depend on mathematical modelling [8]. It used MAP framework to find camera motion with help of smoothness and sparsity. But this method is time consuming and fails on complex details.

Normalization involves adjusting the range of pixel intensity values in an image, removing inconsistencies such as different lighting, contrast and pixel intensity ranges, so that the model focuses on actual features and ignores raw brightness and contrast variations. Various normalization techniques like Batch Normalization [9] and Batch-Instance Normalization (BIN) [10] are employed in image classification. The CNN–Transformer multiscale hybrid architecture employs Layer normalisation within its Transformer modules for image deblurring [11]. Similarly, Layer Normalization is also utilized in DETR [12] while GroupNorm [13] is implemented in FCOS for object detection [14]. Additionally, Instance Normalization has proven effective for style and domain transfer applications [15] [16].

To advance the network performance in image restoration tasks, instance normalization blocks are introduced in HINet [17]. Instance normalization is generally avoided in low-level vision tasks because it may remove important details like contrast and intensity. HINet resolves this issue by adding HIN Blocks which helps model to benefit from normalization without losing important details. HIN Block splits the input features into two halves, one goes through instance normalization which removes information like style and contrast and other half skips normalization that preserves original features. Both halves are then concatenated together. Each stage has an encoder and decoder. In stage one encoder uses multiple HIN blocks to extract features at different scales and decoder reconstructs the image progressively, producing an initial deblurred output. Further ahead stage two takes the output from stage one and refines it using another encoder-decoder with HIN blocks and produces the final sharp image. The network also includes cross-stage feature fusion and a supervised attention module, which helps the two stages share information and combine multi-scale features more efficiently [18].

Though HINet performs well for general image restoration, it treats all image regions equally and does not consider the geometry of the human face. As a result, facial details like eyes, nose, and mouth become extra smooth during normalization. Hence, to resolve this, the paper discusses enhancement over existing HINet. The main contribution of the work is as follows:

1. It improves the original HINet by adding facial structure priors, which help the model to restore blurred face images.
2. It introduces Landmark-Guided Half Instance normalization (LG-HIN), in which facial landmarks help the network to handle features like eyes, nose and mouth carefully.
3. It adds an auxiliary landmark prediction head and uses a joint reconstruction-landmark loss to improve structural consistency

2 Literature Review

Image deblurring is one of the important tasks considering the rapid growth of visual surveillance and authentication systems. Blurred images in security feeds or identity verification can compromise system accuracy, hamper crime investigation highlighting the urgent need for robust deblurring solutions. For facial images, blur can occur due to a variety of environmental and manual factors. Motion blur largely affects the accuracy of pretrained models to classify or recognize elements [19] [20]. Also, sometimes the camera lens is not able to fully converge the light rays on the image sensor, leading to out-of-focus blur. Another type of blur called the Gaussian blur occurs due to the over-smoothening caused by camera lenses or CCTV cameras. These types of blurs negatively affect the recognition accuracy of models [21]. However, in real-life scenarios, facial images face a mix of all these types of blurs, thus necessitating a need for complex deblurring methods.

2.1 Generic methods

Traditional blind image deblurring techniques rely on blur kernel estimation, particularly for motion blur removal. Early approaches estimated blur kernels using hand-crafted priors such as sparsity constraints and total variation. Subsequent improvements focused on enhancing robustness against large motion blur by estimating kernels from salient image structures [22] and by exploiting sparse image representations along with cross-scale self-similarity [23]. These methods improved kernel estimation accuracy and restoration quality.

Most blind kernel deblurring methods primarily address motion blur. However, An et al. proposed a blur-invariant kernel-adaptive network capable of estimating kernels for various types of blurs, rather than being restricted to motion blur alone [24]. Their kernel adaptive autoencoder block jointly encodes features from both the blurred image and the estimated kernel, enabling effective decoding of image and kernel information for recognition tasks. However, these approaches operate on gray-scale kernels, which may limit performance. To address this limitation, Xu et al. introduced an RGB channel-aware blur kernel estimation approach, demonstrating improved PSNR performance compared to gray-scale based methods [25]. While kernel estimation-based methods remain effective for image restoration, their performance degrades when applied to complex, non-uniform real-world blur scenarios.

2.2 Normalization based methods

In deblurring methods, normalization stabilizes neural network training by reducing covariate shift. Early approaches like Batch Normalization in CNNs stabilised training via batch statistics but suffered inter-sample coupling, limiting efficiency in single-image deblurring [26]. Subsequent GAN-based methods addressed this by adopting instance normalization, which generalises better than batch normalization in deblurring generators by normalizing features per individual image across spatial dimensions, eliminating inter-sample dependencies that cause artifacts in single-image tasks. This ensures training stability, faster convergence, and preserved details independent of batch statistics. Further advancements extended instance normalization to discriminators in MIMO-UNet GANs, which normalizes features per image instance across spatial dimensions, minimising batch size dependency [27].

2.3 CNN-based methods

Lim et al. introduced EDSR, a CNN-based enhanced deep residual network for single-image super-resolution. It improves reconstruction quality by using large-scale residual learning and removing batch normalization [28]. Similarly, Chakrabarti et al. introduced a deep learning model where the neural network learns from a paired dataset of blurred and sharp images, thus requiring no need for kernel estimation [29]. Sun et al. extended this idea wherein the network is able to handle spatially varying blur [30]. The model helps handle non-uniform

motion patterns effectively. Various blur-specific methods have been devised to tackle different types of blurs. Some models like DeblurGAN-v2 use Feature Pyramid Network(FPN) with flexible CNN backbones such as MobileNet-v3 and a double-scale RaGAN LS discriminator to provide fast, perceptually sharp image deblurring to improve deblurring performance [31] [32]. The images degraded by Gaussian blur are restored by applying the UNET architecture. This use of encoder-decoder pathways with skip connections effectively captures and restores multi-level image features [33].

2.4 GAN-based methods

Some approaches for the facial motion deblurring include leveraging semantic information and GAN-based two-stage frameworks [34]. Also, coarse-to-fine restoration design using U-NET with robust multi-component loss functions have proven effective. Early GAN-based image restoration methods include ResGAN, a conditional GAN-based image restoration method that uses a ResNet generator to recover images degraded by severe blur or noise [35]. It builds on pix2pix-style cGANs and demonstrates improved PSNR/SSIM on benchmark datasets. Building on the GAN-based restoration framework, Kupyn et al. proposed DeblurGAN, an end-to-end blind motion deblurring framework based on a conditional Wasserstein GAN [36] with perceptual loss, designed to restore sharp images without explicit blur kernel estimation [37]. To address face-specific requirements Meng et al. introduced model, which incorporates the dark channel prior into the loss function for network training [27]. To make it more compatible with a neural network, the L2 norm is adopted rather than the L0 norm. DeblurCGAN advanced this approach specifically for motion blur restoration by combining adversarial loss with content loss based on feature differences extracted by a pretrained VGG19 network to better preserve image textures and details [26]. While DeblurCGAN restored clearer images, this approach was advanced by targeting face recognition, improving both image deblurring quality and recognition accuracy. Zhang et al. categorized DeblurGAN and its v2 variant under adversarial training methods that outperform CNNs in realistic-looking results [38]. Although, GAN based approaches have shown visually appealing results, they still face issues with hallucination and over-smoothing. Also, their high computational cost and large model size limits it's use in low resource surveillance environment.

Thus, HINet provides a lightweight, data-driven deblurring approach. Also, it addresses the loss of structural information caused by full instance normalization in image restoration networks. Instead of normalizing all feature channels, HINet normalizes only half of the channels while keeping the remaining channels unchanged, thus preserving fine-grained details. The network adopts a two-stage progressive restoration framework, where the first stage produces a coarse output and the second stage refines it further. This design has demonstrated strong performance in image deblurring. However, the channel splitting in HINet is fixed and spatially agnostic, limiting its ability to preserve geometric consistency in structure-sensitive applications such as face image deblurring.

3 Methodology

The HINet utilizes an encoder-decoder-based architecture with multi-scale features. In this work, a model is built upon the original HINet framework and extend it by explicitly incorporating facial structural priors. The proposed method introduces landmark-guided feature normalization and auxiliary landmark supervision, enabling the network to better preserve facial geometry while restoring blurred images.

The overall pipeline consists of:

- A two-stage HINet backbone
- Landmark-guided Half Instance Normalization
- A landmark prediction head
- A joint reconstruction landmark optimization strategy

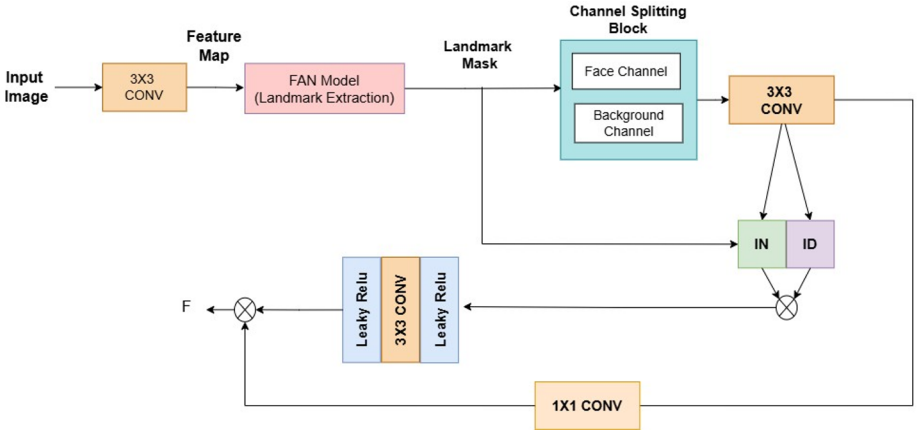

Fig. 1. Architecture of the Landmark-Guided HIN Block

3.1 A two-stage HINet backbone

HINet follows a two-stage U-Net-like encoder-decoder architecture with skip connections that preserve spatial details across multiple resolutions. It consists of the following key components:

- **Cross-Stage Feature Fusion (CSFF):** CSFF transfers intermediate features of encoder and decoder from the first stage to second stage layers, which ensures effective feature reuse and improved reconstruction performance.
- **Supervised Attention Module (SAM):** SAM is located between the two stages, it generates an intermediate restored image and computes an attention map which focuses more on informative regions while suppressing less relevant features.

- **Half Instance Normalization (HIN) Blocks:** HIN blocks are used within the convolutional layers, where feature channels are split into two halves. One half undergoes instance normalization while the other remains unchanged, thus allows the network to benefit from normalization without suppressing important image-specific details.

While this strategy works well for general image restoration, face images contain highly structured regions such as the eyes, nose, and mouth. Excessive normalization on these regions can distort geometry and affect identity preservation. The original HINet performs a random channel split, which may normalize channels important for facial details. To address this limitation, this work introduces a **LG-HIN Block** for selective, landmark-aware normalization.

3.2 Landmark-Guided HIN (LG-HIN) Block

The paper discusses the extension of the HIN block into a LG-HIN Block, where facial landmarks guide which channels should be normalized and which should remain unchanged. Figure 1 illustrates the architecture of the LG-HIN Block, showing how landmark-guided channel splitting and feature normalization are incorporated into the existing HIN block.

Landmark Mask Generation The model uses a Face Alignment Network (FAN) to obtain facial landmarks [39]. FAN detects facial landmarks-key points on a face such as eyes, nose, mouth corners, and jawline. It predicts precise landmark locations even under variations in pose, expression, or occlusion. The predicted landmarks are converted into spatial landmark heatmaps, which are then further aggregated into a binary facial mask $M \in \mathbb{R}^{1\times H\times W}$, where:

$M = 1$ denotes facial landmark-influenced regions (eyes, nose, lips, mouth)
$M = 0$ represents non-facial background regions.

This mask determines which features should be protected from normalization.

Landmark-Guided Channel Splitting Originally, the channel splitting in HINet is random, which does not consider the semantic importance of facial regions, which may lead to distortion of key features such as eyes, nose, and lips. The proposed model improves upon this by introducing Landmark-Guided Feature Splitting.

Using the binary facial mask $M \in \mathbb{R}^{1\times H\times W}$ generated from FAN-predicted facial landmarks, the proposed model splits intermediate features based on semantic regions:

$$F_a = F_{\text{mid}} \odot M \tag{1}$$

$$F_b = F_{\text{mid}} \odot (1 - M) \tag{2}$$

Here, F_a contains facial landmark-influenced regions that are protected from normalization, while F_b corresponds to non-facial regions that can safely undergo IN. The final output is obtained by concatenating the two branches:

$$F_{\text{out}} = [F_a, \text{IN}(F_b)] \tag{3}$$

This modification ensures that critical facial features are preserved, while still allowing normalization to improve the overall feature distribution in non-facial regions. By incorporating semantic guidance into the channel split, the proposed model achieves better restoration quality in key facial regions compared to the original HIN design.

Landmark-Guided Normalization (IN + ID) In the LG-HIN block, the spatial landmark mask M determines which locations inside each channel should be normalized.
For the background branch (F_b):

$$\hat{F}_b = \text{IN}(F_b \odot (1 - M)) \tag{4}$$

meaning that only background regions are normalized.
For the facial region branch (F_a):

$$\hat{F}_a = F_a \tag{5}$$

where facial features are preserved by using an identity (ID) operation to avoid geometric distortion or loss of identity.
The two branches are then concatenated:

$$\hat{F} = [\hat{F}_a, \hat{F}_b] \tag{6}$$

Residual Refinement and Shortcut Connection Following the original HIN block design, the concatenated feature map is passed through:

- A 3×3 convolution,
- Two LeakyReLU activation layers,

to produce residual features $R_{\text{out}} \in \mathbb{R}^{C_{\text{out}} \times H \times W}$

A 1×1 convolution is applied to F_{in} to produce the shortcut feature. The final LG-HIN output is:

$$F_{\text{out}} = R_{\text{out}} + \text{Conv}_{1\times1}(F_{\text{in}}) \tag{7}$$

Overall, the LG-HIN block adapts the strengths of the original HIN block while introducing a semantically meaningful normalization strategy tailored for face deblurring.

3.3 Landmark Prediction Head

To further enforce facial structural consistency during restoration, the proposed model introduces a landmark prediction branch. This branch is attached to the decoder features of the second stage of HINet, using the high-level, context-rich feature maps after the SAM.

The landmark head is designed to predict 68 facial landmark heatmaps, corresponding to standard facial key points (eyes, nose, mouth corners, jawline, etc.). Each heatmap represents a probability distribution over the spatial location of a specific landmark. The head is implemented using lightweight convolutional layers, consisting of two 3×3 convolutions with LeakyReLU activations, followed by a 1×1 convolution and a sigmoid activation to normalize the outputs to $[0, 1]$.

Let F_{stage2} be the second-stage decoder feature map. The landmark prediction is computed as:

$$\begin{aligned}\text{Landmark}_{\text{pred}} = \sigma\Big(&\text{Conv}_{1\times1}\big(\text{LeakyReLU}\big(\text{Conv}_{3\times3}\big(\\ &\text{LeakyReLU}\big(\text{Conv}_{3\times3}(F_{\text{stage2}})\big)\big)\big)\big)\Big)\end{aligned} \tag{8}$$

where σ denotes the sigmoid function.

3.4 Loss Functions

The model predicts the residual image at every stage. This residual is added to the corresponding stage input to generate the restored input. The model uses a combination of two loss functions.

Peak Signal-to-Noise Ratio The model originally uses PSNR, that calculates the pixel-wise difference of the output image and the ground sharp image using mean-squared error. For the two-stage HINet, losses from both stages are averaged.

Landmark Heatmap Loss To ensure that the output images are close to the original ones, and the geometry is intact, proposed model has a supervised landmark prediction using a pixel-wise heatmap loss.

The network predicts a heatmap for each facial landmark, showing where that landmark is most likely to be in the image. The ground-truth heatmaps are created by placing small Gaussian spots at the actual landmark positions. The landmark loss is then calculated as the mean squared error between the predicted and ground-truth heatmaps.

$$\mathcal{L}_{\text{lm}} = \|H_{\text{pred}} - H_{\text{gt}}\|^2 \tag{9}$$

This encourages the network to focus on key facial structures, improving both visual quality and geometric accuracy in the restored images.

Combined Loss Function These loss functions are then combined for the final training.

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{lm}}\mathcal{L}_{\text{lm}} \tag{10}$$

where:

λ_{rec} controls image fidelity
λ_{lm} controls landmark consistency

This multi-loss combination stabilizes training and improves the structural preservation of face features.

4 Results and Discussion

This section discusses the dataset used and its preparation, the experimental setting of the training process, performance analysis and comparison with existing models.

4.1 Dataset Preparation

The model is trained on the Celeb-A dataset available in the free domain on Kaggle which is mostly used for facial image restoration tasks [40]. The dataset contains paired images, which have both blurred images and sharp ground truth face images of celebrities, which are used for supervised training. It contains a total of 1,92,600 facial image pairs covering a diverse range of identities, lighting conditions, and expressions commonly observed in real-world face images. These images are split into train, test and validation, containing 1,54,000, 19300 and 19300 images respectively. Each image has a resolution of 128p x 128p. The variety and scale of the CelebA dataset makes it suitable for training models for image restoration, enabling the proposed method to generalize effectively to a wide range of blurred facial inputs.

4.2 Experimental setup

The model is trained on the dataset with 1×10^5 iterations using PyTorch implementation on an Nvidia GPU to accelerate training. A batch size of 64 and patches of 256 x 256 are used. The learning rate is set to 2×10^{-4} as the default, as used in the original HINet.

4.3 Performance Metrics

To evaluate the efficiency of the proposed deblurring model, both pixel-based and quality metrics are used. PSNR was used to measure the reconstruction accuracy by calculating the difference between the restored image and the ground-truth image. Higher PSNR values indicate better noise suppression and more accurate recovery of image details. SSIM was used to evaluate perceptual similarity

by comparing textures, contrast, and structural information between the two images. SSIM reflects human visual perception and provides a reliable indication of structural preservation. These metrics offer a comprehensive evaluation of both numerical accuracy and perceptual reconstruction quality, allowing for fair comparison with baseline methods.

Fig. 2. Blurred input images (Row 1) and Deblurred output images (Row 2) comparison

4.4 Results and comparison

The proposed model of LG-HINet is tested on the CelebA dataset and its performance is compared with several existing image deblurring models trained on the same dataset. In comparison to DeblurGAN and DeblurGAN-v2, which introduced adversarial learning and perceptual loss to enhance visual realism, the proposed LG-HINet adopts a reconstruction-oriented strategy. These models focus more on perceptual quality over pixel-wise loss as a result of adversarial optimization that does not directly minimize reconstruction error which then leads to suboptimal PSNR. In contrast, LG-HINet eliminates adversarial instability by leveraging hierarchical interaction and gradual improvement to reduce mean squared error, resulting in better PSNR. ResGAN uses residual learning within GAN framework to make training stable and improve gradient flow. Similarly, SIUN utilized a U-Net style encoder-decoder with skip connections to preserve structural information. These models' ability to recover fine-grained and spatially varying blur patterns is restricted as they lack cross-level interaction and structural priors. LG-HINet addresses these limitations by introducing a dual-stage restoration mechanism along with Hierarchical Interaction Modules (HIM), which enables bidirectional feature exchange across encoder and decoder stages. This hierarchical design improves feature reuse and semantic consistency, reducing information loss during deblurring.

In Figure 2, row 1 contains blurred images given as input to the model and row 2 contains deblurred output images obtained. The difference shows the qualitative results and effectiveness of the proposed model in restoring blurred face images. Visually, the restored images exhibit enhanced sharpness, improved contrast, and more accurate reconstruction of key facial attributes such as the eyes, nose, and lips while preserving the overall facial geometry, compared to other similar methods.

Table 1 shows the comparative study of various existing models on quantitative performance metrics, showing improvements over existing approaches. The baseline HINet when trained on CelebA achieves a PSNR of 24.16 and a SSIM of 0.86, whereas the proposed model improves the results by achieving a PSNR of 26.2 and SSIM of 0.91. This highlights the effectiveness of the modifications introduced in the proposed approach.

The improvement in PSNR is mainly due to the integration of landmark based loss functions, which guide the network to reduce reconstruction errors in important facial regions. Similarly, the increase in SSIM is due to the use of facial landmark priors, which help preserve structural consistency and spatial relationships. By focusing on edges and structural features near the landmark regions, the model enhances retention of facial geometry and textures, which contributes to improved structural similarity.

Also, the model demonstrates stable and progressive learning behaviour, improving performance with increasing training iterations. Overall, both quantitative and qualitative evaluations show that the proposed method significantly outperforms prior approaches in face image deblurring.

Table 1. Quantitative comparison of different models

Model	**PSNR (dB)**	**SSIM**
DeblurGAN [37]	23.48	0.75
DeblurGAN-v2 [31]	21.10	0.75
ResGAN [35]	20.25	0.44
SIUN [7]	22.16	0.84
EDSR [28]	22.92	0.76
HINet (Baseline)	24.16	0.86
LG-HINet	**26.2**	**0.91**

5 Conclusion and Future Scope

Usual image deblurring methods often lead to loss of originality and hamper recognition. So, the proposed work presents a lightweight CNN based facial image deblurring framework with landmark guided enhancement to HINet for face image deblurring. The model introduces landmark-based channel splitting and landmark-aware loss functions during training that improves feature preservation in critical facial areas.

The proposed modifications enhance both pixel-level accuracy and structural similarity in restored images. It achieves a PSNR of 26.2 and SSIM of 0.91 demonstrating high deblurring power of the model. The model also provides lower computational cost making it suitable for edge based application. Thus, the model outperforms others making it suitable for intelligent surveillance system supporting sustainable image deblurring.

Although the model shows promising results yet there are few directions for future research. First, the current model uses FAN pretrained model for landmark detection which can later be improved to make an end-to-end joint framework to improve robustness. Second, the deblurring network can be integrated with face recognition pipelines to improve recognition performance as it holds great potential for security related applications.

References

1. R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE Trans. Image Process.*, vol. 29, pp. 6251–6263, 2020.
2. Z. Shen, W. Lai, T. Huang, and M. H. Yang, "Exploiting Semantics for Face Image Deblurring," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1829–1846, 2020. doi: 10.1007/s11263-019-01288-9.
3. H. Tian, L. Sun, X. Dong, B. Lu, H. Qin, L. Zhang, and W. Li, "A Modelling Method for Face Image Deblurring," *2021 International Conference*, 2021.
4. S. Zhang, A. Zhen, and R. L. Stevenson, "GAN-Based Image Deblurring Using Dark Channel Prior," *University of Notre Dame, presented at the International Conference*, 2019.
5. H. Wang and S. Wang, "Low-Resolution Face Recognition Enhanced by High-Resolution Facial Images," in *Proc. 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, Waikoloa Beach, HI, USA, 2023, pp. 1–8. doi: 10.1109/FG57933.2023.10042552.
6. S. H. Jung, T. B. Lee, and Y. S. Heo, "Deep Feature Prior Guided Face Deblurring," *Ajou University, South Korea.*
7. M. Ye, D. Lyu, and G. Chen, "Scale-iterative upscaling network for image deblurring," *IEEE Access*, vol. 8, pp. 18316–18325, 2020.
8. T. Lokesh Kumar, *Deep Unsupervised Single Image Deblurring*, Project Report, IIT Madras, 2021.
9. F. Albluwi, V. A. Krylov, and R. Dahyot, "Image Deblurring and Super-Resolution Using Deep Convolutional Neural Networks," in *Proc. 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, Aalborg, Denmark, 2018, pp. 1–6. doi: 10.1109/MLSP.2018.8516983.

10. J. Li, D. Lin, H. Chen, H. Liu, L. Wan, and W. Feng, "Causal Inference Via Style Bias Deconfounding for Domain Generalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. doi: 10.1109/TPAMI.2026.3652609.
11. Q. Zhao, H. Yang, D. Zhou, and J. Cao, "Rethinking Image Deblurring via CNN-Transformer Multiscale Hybrid Architecture," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, Art. no. 5002415, pp. 1–15, 2023. doi: 10.1109/TIM.2022.3230482.
12. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," *European Conference on Computer Vision*, Springer, 2020, pp. 213–229.
13. T. Germer, T. Uelwer, and S. Harmeling, "Deblurring Photographs of Characters Using Deep Neural Networks," *arXiv:2205.15053 [cs.CV]*, 2022.
14. Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9627–9636.
15. X. Wang, W. Wang, S. Yang, and J. Liu, "CLAST: Contrastive Learning for Arbitrary Style Transfer," *IEEE Transactions on Image Processing*, vol. 31, pp. 6761–6772, 2022. doi: 10.1109/TIP.2022.3215899.
16. N. Halagatti *et al.*, "Efficient Semantic Image Editing with Feature Shared Spatially-Adaptive Normalization," in *Proc. 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, Silchar, India, 2023, pp. 1–7. doi: 10.1109/ISACC56298.2023.10083951.
17. L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen, "HINet: Half Instance Normalization Network for Image Restoration," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2021, pp. 182–192.
18. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 14821–14831.
19. K. Knežević, E. Mandić, R. Petrović, and B. Stojanović, "Blur and Motion Blur Influence on Face Recognition Performance," in *Proceedings of the 14th Symposium on Neural Networks and Applications (NEUREL)*, Belgrade, Serbia, 2018, pp. 1–5, doi: 10.1109/NEUREL.2018.8587028.
20. I. Vasiljević, A. Chakrabarti, and G. Shakhnarovich, "Examining the Impact of Blur on Recognition by Convolutional Networks," *arXiv preprint arXiv:1611.05760*, 2016.
21. F. Hua et al., "Impact of Out-of-focus Blur on Face Recognition Performance Based on Modular Transfer Function," *IEEE BTAS*, 2010.
22. J. Pan, R. Liu, Z. Su, and X. Gu, "Kernel estimation from salient structure for robust motion deblurring," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1156–1170, 2013.
23. J. Yu, Z. Chang, and C. Xiao, "Blur Kernel Estimation Using Sparse Representation and Cross-Scale Self-Similarity," *Multimedia Tools and Applications*.
24. S. An, H. Roh, and M. Kang, "Blur Invariant Kernel-Adaptive Network for Single Image Blind Deblurring," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, China, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428294.
25. X. Xu, H. Liu, Y. Li, and Y. Zhou, "Image deblurring with blur kernel estimation in RGB channels," in *Proceedings of the IEEE International Conference on Digital Signal Processing (DSP)*, Beijing, China, 2016, pp. 681–684, doi: 10.1109/ICDSP.2016.7868645.

26. Z. Su, H. Zhang, Z. Liu, and X. Zhu, "Image Deblurring Algorithm by Using Conditional Generative Adversarial Network," in *Proceedings of the 40th Chinese Control Conference (CCC)*, Shanghai, China, 2021, pp. 8128–8133, doi: 10.23919/CCC52363.2021.9549873.
27. Q. Meng and H. Zhang, "GAN-Based Encoder-Decoder Method for Image Deblurring," *RAIIE*, IEEE, 2025. DOI: 10.1109/RAIIE65740.2025.11140515.
28. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
29. A. Chakrabarti, "A Neural Approach to Blind Motion Deblurring," *Toyota Technological Institute at Chicago*, supp_results.pdf, accessed 7 Nov. 2025.
30. 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 (CVPR)*, Boston, MA, USA, 2015, pp. 769–777, doi: 10.1109/CVPR.2015.7298677.
31. O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, 2019, pp. 8877–8886, doi: 10.1109/ICCV.2019.00897.
32. S. Huang and W. Fu, "Deblur GAN-v2 Based Mechanical Pipeline Defuzzification Algorithm," in *Proc. 2nd Int. Conf. Frontiers Electron., Inf. Comput. Technol. (ICFEICT)*, Wuhan, China, 2022, pp. 416–420, doi: 10.1109/ICFEICT57213.2022.00080.
33. A. J. Ajwad and S. T. S. Rafid, "Enhancing Facial Image Clarity: Deblurring Gaussian Blur with UNET++ Architecture," in *Proc. Int. Conf. Computer and Information Technology (ICCIT)*, IEEE, 2023.
34. K. Xiao and Z. Pan, "A Facial Motion Deblurring Algorithm Based on Semantics and GAN," *TCS*, IEEE, 2023.
35. M. Wang, H. Li, and F. Li, "Generative adversarial network based on ResNet for conditional image restoration," *arXiv:1707.04881*, 2017.
36. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.
37. O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8183–8192.
38. K. Zhang, W. Ren, W. Luo, W.-S. Lai, B. Stenger, M.-H. Yang, and H. Li, "Deep Image Deblurring: A Survey," *International Journal of Computer Vision*, 2022.
39. A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 1021–1030HI
40. Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.