Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.5359
Published by: Zeus Press

Advances in Deep Learning-Based Face Deblurring Methods

Haochen Wang*

College of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

**Corresponding author: Haochen Wang.*

Abstract

The task of face deblurring is to restore a clear, realistic and identity-preserving high-quality face image from a given blurred face image. This paper systematically expounds the methods of face deblurring based on deep learning from five stages: traditional methods based on physical models, early deep learning methods based on end-to-end regression, methods based on generative adversarial networks (GAN) and prior embedding, methods based on the Transformer architecture, and methods based on diffusion models, following the development of technology. By sorting out the core ideas, key technologies and representative works of each stage, it reveals the clear evolution of face deblurring research from "pixel-level image restoration" to "identity-constrained generative modeling". In addition, it introduces the commonly used datasets for face deblurring, and makes predictions and prospects for the problems to be solved and future research directions in face deblurring research. Face deblurring is a research hotspot in the field of computer vision, and more high-quality algorithms will be proposed in the future, and it is developing in a more diversified direction.

Keywords

face deblurring, deep learning, identity preservation, diffusion model

1. Introduction

Face deblurring is an important research topic in the field of computer vision, aiming to recover clear and vivid facial details from blurry facial images. This technology has extensive application values in scenarios such as security surveillance, mobile photography, photo restoration, and the preprocessing of face recognition systems. Different from ordinary image deblurring, face images possess highly structured prior knowledge and identity sensitivity, which impose higher requirements on the repair ability of algorithms. Not only is it necessary to remove the blur artifacts, but also to ensure that the restored face and the original identity can be kept consistent.

Face deblurring methods have evolved from traditional mathematical optimization approaches to deep learning-driven paradigms. Early methods physical-model-based estimating blur kernels and then performing deconvolution to restore images are however restricted by the expressive ability of manual priors and struggled to handle complex dynamic scenes. Since 2014, deep learning methods represented by convolutional neural networks (CNN) have replaced explicit blur kernel estimation with end-to-end mapping, significantly improving restoration quality. Subsequently the introduction of GANs expands the optimization objective from

pixel precision to perceptual realism meanwhile the embedding of facial specific priors makes the algorithm start to focus on the core issue of identity preservation. Recently, Transformers and diffusion models have further advanced the field. In general, based on this technical evolution, this paper categorizes face deblurring methods into five stages: the era of physical models, the early stage of deep learning, the era of generative priors and prior embeddings, the era empowered by the Transformer architecture and the diffusion generation era.

The evolution of these five stages, essentially speaking, represents a dual transformation of the a priori forms and the optimization objectives. Initially the innate form was the early manual design then gradually moved towards the data-driven direction and finally became the implicit learning of the generative model. As for the optimization objective initially it mainly focused on pixel precision then attached more importance to perceptual quality and finally also raised the priority of identity constraints.

Within this review framework, this paper systematically reviews and summarizes the research progress and current situation of the deep-learning-driven facial deblurring algorithms. It elaborates in detail the commonly used datasets and evaluation metrics in this field as well as the experimental results of various algorithms on mainstream datasets. Moreover, it analyzes and predicts the research directions and development trends that are future-oriented for face deblurring.

2. Classification and Evolution of Face Deblurring Methods

2.1 Phase 1: The Era of Physical Models

The main method is blind deconvolution. It estimates both the blur kernel and the clear image from a single blurred input by modeling $B = I \otimes k + n$. With prior constraints on image and kernel, it optimizes posterior probability to estimate k, then recovers I via non-blind deconvolution. Representative methods include variational Bayesian blind deconvolution and hyper-Laplacian prior-based non-blind deconvolution.

The variational Bayesian blind deconvolution method manually selects image patches, fits the gradient distribution with a Gaussian mixture and blur kernel shape with an exponential distribution, approximates the posterior via variational Bayesian, estimates the kernel in a coarse-to-fine manner, and restores the image using Richardson-Lucy. It was the first to introduce variational Bayesian into blind deconvolution, avoiding invalid MAP solutions and laying a foundation for modern methods. However, it is computationally expensive, requires manual patch selection, is noise-sensitive, and prone to ringing artifacts [1].

The hyper-Laplacian prior-based non-blind deconvolution method fits natural image gradients more accurately using a hyper-Laplacian distribution and applies an alternating minimization scheme to decompose the problem into two easier sub-problems, greatly improving efficiency. It reduces deconvolution time for megapixel images from 20 minutes to about 3 seconds, providing an efficient solution. However, it relies on accurate blur kernel input, only handles uniform blur, and cannot address spatially varying blur—a common limitation of traditional optimization methods [2].

2.2 Phase Two: The Early Stage of Deep Learning

The main method of this stage is end-to-end CNN regression, which directly learns the mapping from blurry to clear images using deep networks. Typically, a decoder-encoder structure takes the blurry image as input and outputs the clear image, trained on paired data with pixel-level loss. Multi-scale or recursive connections are introduced to handle large-scale blur. Representative methods include the multi-scale CNN, the scale-recurrent network, and SRGAN.

The multi-scale CNN uses a three-scale encoder-decoder structure, where outputs from lower scales guide higher scales, and a multi-scale loss supervises each level. It also introduces the GOPRO dataset with real blur-sharp pairs. This method is the first to achieve end-to-end deblurring for dynamic scenes, initiating the multi-scale deep deblurring paradigm and addressing spatially varying blur. However, its outputs tend to be overly smooth and lack realistic high-frequency textures, a limitation later verified in perceptual loss and GAN-based approaches [3].

Built upon the multi-scale CNN, the scale-recurrent network introduces a recursive structure with parameter sharing across scales, reducing the parameter count to about 1/4 of its predecessor while maintaining multi-scale modeling. It has become a benchmark method for multi-scale deblurring. However, it still relies on pixel-level losses, limiting perceptual quality, and struggles with large occlusions or extreme blur [4].

SRGAN introduces perceptual loss, replacing pixel-level MSE with feature differences from a pre-trained VGG network, and is the first to apply GAN to image restoration, enhancing realism through adversarial training. It marks a paradigm shift from pixel accuracy to perceptual quality, laying the foundation for subsequent GAN-based deblurring. However, it is designed for super-resolution, with limited effect on deblurring, and suffers from GAN training instability, a widely discussed issue in later studies [5].

2.3 Phase 3: The Era of Generative Priors and Prior Embeddings

This stage can be divided into two periods. The first period is general GAN deblurring. It uses a conditional GAN with generator G mapping blurry to clear images and discriminator D distinguishing real from generated results. The loss combines adversarial, content, and perceptual terms, with PatchGAN improving local detail realism. Representative methods include the DeblurGAN series.

DeblurGAN which is the very first one that applies GAN to image deblurring is present. It uses a PatchGAN discriminator on 70×70 patches, also introduces perceptual loss, and then uses WGAN-GP training to ensure stability. This approach which significantly enhances visual realism and achieves fast inference that is suitable for near-real-time applications. However, its capacity to deal with complex motion blur is quite limited and it also continuously generates unnatural texture flaws and so forth [6].

DeblurGAN-v2 which has adopted a feature pyramid network as the generator's backbone is to actually integrate multi-scale features. It supports various backbones, balances speed and accuracy in different scenarios, and also introduces global and local discriminators. While maintaining high-quality generation, the inference speed is greatly increased, and it becomes a common benchmark in engineering applications. However, there still exists the situation of unstable training of GANs as well as insufficient retention of identities [7].

The second period focuses on face-guided priors, which fall into three categories. The first uses geometric constraints such as 3DMM or keypoint heatmaps to ensure facial structure recovery. The second is identity feature alignment: this route extracts identity features through a pre-trained face recognition network and constructs an identity loss function using cosine similarity, which effectively raises identity consistency from around 0.7 to over 0.85. Nevertheless, it heavily relies on the generalization performance of the recognition model and may overly restrict natural changes in facial expressions and postures. The third leverages pre-trained StyleGAN priors via latent space search/optimization or CNN feature fusion. Representative works include PULSE, GFP-GAN, and CodeFormer.

The task of face deblurring is redefined by PULSE into a problem of potential space search. By randomly sampling or optimizing the latent vectors in the StyleGAN latent space the generated images are downsampled to match the input blurred images instead of directly repairing the blurred images. This method fundamentally changes the traditional deblurring solutions and can extract high-fidelity details from low-resolution blurred inputs yet there exists the problem that the generated results do not conform to the real identity [8].

GFP-GAN combines U-Net with a pre-trained StyleGAN. It fuses CNN features with the intermediate features of StyleGAN through a spatial feature transformation layer, preserving identity and restoring details. It also uses facial component loss to enhance key regions like eyes and mouths. At that moment, GFP-GAN, which had attained leading performance in terms of identity similarity and perceptual quality, became a landmark work in facial restoration. However, it is sensitive to input image quality and suffers identity drift under large poses or severe blurring [9].

A Transformer structure that has a codebook prediction mechanism and is developed by CodeFormer carries out discrete representation learning on the latent codes of the blurry faces. It gets the pre-trained codebook to acquire high-quality features and can flexibly control the repair intensity through weight adjustment. As the first adjustable face restoration method, it balances between identity preservation and perceptual realism. Its performance to a large extent depends on the coverage situation of the codebook and its adaptability to rare identities and extreme postures remains insufficient [10].

2.4 Phase 4: The Era Empowered by the Transformer Architecture

These methods that rely on self-attention to carry out global modeling have surmounted the limitation of the local receptive field of the CNN and can also seize the dependencies among the long-distance pixels. The self-attention module is embedded in the encoder-decoder structure so as to retain multi-scale features and global context, and at the same time, the window/channel attention, which reduces the computational cost of high-resolution face images. For instance, there are representative methods like Restormer, SwinIR, and Uformer.

A multi-depth convolutional head transposed attention (MDTA) is put forward by Restormer, which calculates attention along the channel dimension and reduces the complexity from $O(n^2)$ to $O(n)$. It also uses a gated feedforward network and stacks MDTA modules in the encoder-decoder path for multi-scale global-local modeling. This approach, which has achieved state-of-the-art performance in image deblurring, deraining and denoising and has significantly higher PSNR/SSIM, demonstrates Transformer effectiveness in low-level vision. However, its perception measurement and identity similarity are slightly lower than those of the current GANs and diffusion models [11].

SwinIR is based on the Swin Transformer and adopts a three-stage structure of shallow feature extraction + deep feature extraction + high-quality image reconstruction. The RSTB module internally achieves cross-window information interaction through window division and shifting, balancing local and global modeling. This method is the first to apply the shifted window attention system to image reconstruction and becomes a pioneer of the Transformer for image restoration. However, it has a high computational cost and has limited recovery effect for input with high degree of blurriness [12].

Uformer embeds the LeWin Transformer module in the encoder and decoder paths of U-Net, performing self-attention calculations on the downsampled feature maps; it fuses multi-scale features through skip connections, balancing detail preservation and global modeling. This method enhances the global perception ability while maintaining the multi-scale advantage of U-Net, and performs well on multiple deblurring benchmarks. However, its ability to handle large-area blurring is still inferior to the pure Transformer structure [13].

2.5 Phase Five: The Diffusion Generation Era

The main method is the denoising diffusion probabilistic model. Slowly it adds noise to the clear image until it turns into pure noise and then learns to reverse this process to restore the image. For deblurring, the blurred image acts as a condition. The forward process is a fixed Markov chain which will be added with Gaussian noise; the reverse process is to learn a parameterized Markov chain that continuously iterates for denoising in order to produce high-quality results. Representative methods include SR3, DiffBFR, ID-Blau and TD-BFR.

SR3 captures low-resolution images, upsamples them, and then uses these as conditions to direct the reverse denoising procedure of the diffusion model. Using a U-Net backbone to predict noise at each iteration after thousands of steps high-resolution details are generated. This approach attains top performance in image super-resolution, showing the potential of diffusion models in image restoration. However, its reasoning speed is extremely slow, which restricts practical applications and has become the core optimization direction for subsequent diffusion model research [14].

DiffBFR is a diffusion model for blind face restoration. It incorporates an identity restoration module that maintains identity consistency through a pre-trained face recognition network, together with a texture enhancement module to refine local facial details. Multi-scale condition injection improves robustness to complex blurs. This method outperforms GAN-based approaches on several real-world blurred datasets, yielding identity similarity scores above 0.94 and achieving leading perceptual quality. Nonetheless, its inference efficiency remains a critical limitation, and restoration results for extremely blurred face images still require further improvement [15].

ID-Blau devises an identity decoupling diffusion model, and the identity mask learning strategy is employed to separate the identity features and the texture features. It can generate multiple plausible outputs from the same ambiguous input, realizing result diversity, and also introduces uncertainty estimation to focus on severely ambiguous regions. As the first diffusion model supporting diverse generation, it is quite compatible with the ill-posed nature of this problem. However, its mechanism for diversity control is still not entirely perfect and some outputs may deviate from the original identity [16].

TD-BFR addresses slow diffusion inference by introducing a low-resolution sampling strategy that reduces iterative steps, along with an adaptive degradation module that adjusts the sampling trajectory based on blur severity. These improvements boost inference speed by 4.75 times while maintaining competitive generation quality, marking an important step toward practical deployment of diffusion-based face restoration. However, its ability to be extended to different fuzzy types and to maintain identity still needs to be further verified, for the latter is slightly lagging behind the original diffusion model [17].

3. Data set

In the research of face deblurring, many datasets have been used, mainly divided into synthetic datasets and real-blurred datasets. Synthetic datasets generate precise pairs of blurred and clear images by convolving clear images with predefined blur kernels or by capturing precise blurred images through high-speed cameras that record real motion trajectories, providing pixel-level ground truth for supervised learning. Real-blurred datasets are collected from actual scenes, with the blurring formed naturally by factors such as camera shake, object movement, or blur, which are closer to application requirements but usually lack corresponding clear reference images, and are mainly used to test the generalization ability of the model. The widely used datasets in face deblurring research cover general deblurring datasets, face-specific datasets, and real-blurred datasets, providing a basis for training and testing algorithms at different stages. Detailed information about the relevant datasets is shown in Table 1.

Table 1: Common Datasets for Face Deblurring

Dataset name	Year	Number of images	Resolution	Fuzzy type	Features and Applications
GOPRO	2017	3214 pairs (for training) / 1111 pairs (for testing)	1280×720	Motion blur	Common deblurring benchmark; dynamic scenes; limited faces, used for pre-training
HIDE	2018	8422 pairs	1280×720	Motion blur+Character movement	Human/face scenes; annotated bounding boxes
RealBlur	2019	3758 pairs (for training) / 980 pairs (for testing)	Various	Real camera shake	Real captured blur-sharp pairs; RealBlur-J/R subsets
CelebA-HQ	2018	30,000 high-definition human faces	1024×1024	/	High-quality face dataset; synthetic blur for train/test pairs
FFHQ	2019	70,000 high-definition human faces	1024×1024	/	Diverse faces; widely used for generative priors
Wider Face	2016	32,203 images (393,703 faces)	Various	Natural scene is blurry.	Face detection dataset; real blurred faces; tests generalization
LSDIR	2022	5000 pairs (for training) / 1000 pairs (for testing)	Various	Real + Synthetic Blur	Large-scale real image deblurring; multiple degradations; includes faces
UMDFaces	2017	22,075 video frames	Various	Real motion / Blur due to defocus	Video face dataset; real-world blur; scene testing
Face-Synthetics	2021	100,000 synthetic human faces	1024×1024	Can be customized	Fully synthetic; arbitrary blur kernels; large-scale training

4. Conclusion

This article makes a comprehensive and systematic review of the development process of the face deblurring method that is based on deep learning. Based on the evolution of the technological paradigm, it is divided into five phases: the era of physical models, the early stage of deep learning, the era of generative priors and prior embeddings, the era empowered by the Transformer architecture and the diffusion generation era. It summarizes the core ideas key technologies and representative works of each phase clearly presenting the trajectory of face deblurring evolving from “pixel-level image restoration” to “identity-constrained

generative modeling". Future research will continue to deepen in many directions such as efficiency optimization, multimodal fusion, controllable generation, and ethical norms, thereby promoting this technology to move towards a more reliable and more practical direction.

References

[1] Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., & Freeman, W. T. (2006). Removing camera shake from a single photograph. In Acm Siggraph 2006 Papers (pp. 787-794).

[2] Krishnan, D., & Fergus, R. (2009). Fast image deconvolution using hyper-Laplacian priors. Advances in neural information processing systems, 22.

[3] Nah, S., Hyun Kim, T., & Mu Lee, K. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3883-3891).

[4] Tao, X., Gao, H., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8174-8182).

[5] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[6] Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8183-8192).

[7] Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2019). Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8878-8887).

[8] Menon, S., Damian, A., Hu, S., Ravi, N., & Rudin, C. (2020). Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 2437-2445).

[9] Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9168-9178).

[10] Zhou, S., Chan, K., Li, C., & Loy, C. C. (2022). Towards robust blind face restoration with codebook lookup transformer. Advances in Neural Information Processing Systems, 35, 30599-30611.

[11] Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., & Yang, M. H. (2022). Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5728-5739).

[12] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).

[13] Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., & Li, H. (2022). Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 17683-17693).

[14] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M. (2022). Image super-resolution via iterative refinement. IEEE transactions on pattern analysis and machine intelligence, 45(4), 4713-4726.

[15] Qiu, X., Han, C., Zhang, Z., Li, B., Guo, T., & Nie, X. (2023, October). Diffbfr: Bootstrapping diffusion model for blind face restoration. In Proceedings of the 31st ACM international conference on multimedia (pp. 7785-7795).

[16] Wu, J. H., Tsai, F. J., Peng, Y. T., Tsai, C. C., Lin, C. W., & Lin, Y. Y. (2024). Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 25847-25856).

[17] Zhang, Z., Gao, X., Wang, Z., & Zhang, X. (2025). TD-BFR: Truncated diffusion model for efficient blind face restoration. arXiv preprint arXiv:2503.20537.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights