A Lightweight Hybrid CNN-Transformer Model for Smoke Removal in Laparoscopic Surgery Images

Mohammed Ashlab N 23MIA1081
School of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
mohammedashlab.n2023@vitstudent.ac.in

Dhinesh Kumar M 23MIA1102
School of Computer Science & Engieering
Vellore Institute of Technology
Chennai, India
dhineshkumar.m2023@vitstudent.ac.in

Srishreyan S 23MIA1088
School of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
srishreyan.s2023@vitstudent.ac.in

Abstract— Laparoscopic surgeries are highly dependent on visual clarity, but the generation of surgical smoke during procedures such as cauterization and ablation often obscures the surgeon's view and degrades image quality [14]. To overcome this issue, we propose a Progressive Frequency-Aware Network (PFAN)—a lightweight hybrid deep learning architecture that integrates Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) within a Generative Adversarial Network (GAN) framework [6], [5]. The proposed model progressively extracts and fuses high-frequency local and low-frequency global features in the frequency domain to restore clear, artifact-free laparoscopic images. PFAN introduces two novel components: the Multi-scale Bottleneck-Inverting (MBI) Block for multi-scale high-frequency extraction and the Locally-Enhanced Axial Attention Transformer (LAT) Block for global low-frequency feature representation. The model is trained using synthetic smoke images generated via the Blender engine to simulate realistic laparoscopic smoke [3]. Experimental evaluation on the Cholec80 dataset [28] demonstrates that PFAN achieves superior quantitative results in PSNR, SSIM, and CIEDE2000 metrics compared to state-of-the-art approaches such as Pix2Pix [13], CycleGAN [19], and DCP [7], while maintaining only 629K parameters, enabling real-time deployment on resource-constrained surgical systems.

Keywords— Laparoscopic image desmoking, deep learning, GAN, CNN, Vision Transformer, frequency-aware learning, lightweight architecture, medical image enhancement.

I. INTRODUCTION AND MOTIVATION

Laparoscopic surgery has become one of the most common minimally invasive procedures in modern medicine, offering advantages such as smaller incisions, faster recovery, and reduced postoperative pain [14]. However, surgical smoke generated during tissue dissection or cauterization significantly impairs the visibility of critical anatomical structures, posing a risk to surgical accuracy and patient safety.

To mitigate this, hardware-based smoke evacuation systems are typically used, but their high cost and limited adaptability make them unsuitable for widespread adoption [27]. Consequently, image-based desmoking algorithms have emerged as an attractive alternative, capable of enhancing image clarity through computational processing.

Traditional approaches, such as the Dark Channel Prior (DCP) [7], Retinex-based enhancement [22], and wavelet transform algorithms [24], rely on atmospheric scattering models to estimate and remove haze-like smoke. However, these techniques often produce visual artifacts and distort colors due to their assumptions of uniform smoke distribution, which do not hold true in surgical environments [29].

With the advent of deep learning, methods based on Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) [6] have achieved remarkable progress in tasks such as dehazing, denoising, and image-to-image translation [13], [19]. Despite their success, CNN-based models primarily capture local spatial features and struggle with global contextual dependencies, which are crucial for understanding the overall structure of laparoscopic scenes. In contrast, Vision Transformers (ViTs) excel in modeling long-range dependencies [5], but their computational complexity limits their use in real-time medical imaging applications [17].

To address these challenges, we implement the Progressive Frequency-Aware Network (PFAN)—a hybrid CNN–Transformer architecture designed for effective laparoscopic image desmoking. PFAN operates in the frequency domain, progressively extracting both high-frequency (local detail) and low-frequency (global structure) features through a Generative Adversarial Network (GAN) framework [6]. The architecture consists of two main modules:

1. a Multi-scale Bottleneck-Inverting (MBI) Block for multi-scale high-frequency feature extraction [12], and

2. a Locally-Enhanced Axial Attention Transformer (LAT) Block for efficient global low-frequency information modeling [17].

Unlike traditional methods, PFAN achieves high desmoking performance while maintaining only 629K parameters, making it lightweight and suitable for resource-constrained surgical devices. Evaluations on the Cholec80 dataset [28] demonstrate that PFAN outperforms previous state-of-the-art models, achieving higher PSNR, SSIM, and CIEDE2000 scores, and delivering visually clearer and color-consistent results.

The primary contributions of this work can be summarized as follows:

We develop a lightweight hybrid CNN-Transformer framework for laparoscopic image desmoking within a GAN structure.

We introduce the MBI and LAT Blocks, designed to extract progressive frequency-domain features efficiently.

We employ a Blender-based synthetic smoke generation method for creating diverse and realistic training data [3].

We demonstrate the superiority of PFAN over existing traditional and deep learning-based desmoking methods [7], [13], [19], [29].

II. RELATED WORK

A. Image Dehazing and Desmoking Methods

Image dehazing has been one of the most studied topics within computer vision, and early approaches were based on physical models of atmospheric scattering. The so-called Dark Channel Prior (DCP) from He et al. [1] became a seminal work in this domain, with its idea being that most local patches in haze-free outdoor images contain some pixels with very low intensity in at least one color channel. Demonstrating distinct advancement over the results of earlier techniques, this prior-based approach nevertheless suffered from computationally intensive operations and scenes that violated the assumption of the dark channel.

Several variants were built on top of these physical models to overcome the limitations of DCP. For example, Zhu et al. [2] present a color attenuation prior for developing a linear model in scene depth based on the difference between brightness and saturation. Berman et al. [3] propose non-local prior based methods that assume haze-free images are dominated by a few hundred distinct colors forming tight clusters in RGB space. These demonstrated some promise for dehazing in outdoor settings but had limited applicability to medical imaging, particularly in laparoscopic procedures, due to the unique nature of surgical smoke.

B. Deep Learning Approaches for Image Restoration

Deep learning heralded a new era for image restoration problems by showing that CNN-based techniques far outperform traditional methods. Cai et al. [4] initiated the application of CNNs in the imaging dehazing area by proposing DehazeNet for estimating the transmission map directly from a hazy image. This work marked the beginning of the paradigm shift from hand-crafted features to learned representations.

Later architectures came with more enhanced designs. Ren et al. [5] suggested a multi-scale deep neural network that jointly estimates transmission maps and atmospheric light. Li et al. [6] proposed AOD-Net, an All-in-One Dehazing Network that reformulated the atmospheric scattering model into a unified framework; this has given rise to end-to-end learning without explicit estimation of intermediate parameters. These techniques proved that learning all components of dehazing jointly performs better than sequential estimation.

With the introduction of encoder-decoder architectures, restoration capabilities were further extended. Qu et al. [7] developed an improved pix2pix network with GANs to enhance the quality of the dehazed images. Zhang et al. [8] proposed the Densely Connected Pyramid Dehazing Network, which utilized dense connections so that there was better feature reusability and propagation of gradients. However, most general-purpose dehazing networks fail to clearly capture specific characteristics related to surgical smoke in laparoscopic environments.

C. Frequency Domain Analysis in Image Processing

Frequency domain analysis has long been recognized as a powerful tool for the understanding and manipulation of image content. Conventional approaches, such as Fourier Transform and DCT, have been widely applied in image compression and filtering [9]. Very recently, there has been active work exploring the integration of frequency domain information into deep learning architectures.

Chi et al. [10] showed the potential for neural network performance improvements in several vision tasks by incorporating frequency domain knowledge. They found that distinct components in the frequency domain capture complementary information in image structure. Then, Xu et al. [11] went on to propose the dual-domain network, which could process spatial and frequency information in parallel and achieved state-of-the-art image denoising results.

Frequency domain approaches have shown special promise in the context of image restoration. Fuoli et al. [12] have presented frequency separation techniques for video super-resolution, and demonstrated that processing different frequency bands separately preserves fine details while removing noise. On the other hand, progressive frequency-aware processing applied to medical image desmoking remains largely under-explored.

D. Medical Image Enhancement and Laparoscopic Vision

Medical image enhancement differs from natural image processing in that it has very strict requirements on both diagnostic accuracy and real-time performance. The early work of laparoscopic image enhancement was focused on color correction and illumination normalization [13]. Tchoulack et al. [14] have proposed methods tailored for laparoscopic image processing, considering issues such as specular reflections and non-uniform illumination.

Recent advances have applied deep learning to surgical vision tasks. Luo et al. [15] proposed a CNN-based approach for laparoscopic image desmoking, showing the feasibility of learning-based approaches in surgical environments. Bolkar et al. [16], on the other hand, presented a smoke removal system using encoder-decoder architectures with skip connections that improved the visibility in laparoscopic videos.

The LID dataset introduced by Luo et al. [15] is the first large-scale benchmark for evaluating desmoking algorithms in surgical contexts. Indeed, this dataset has allowed for the first systematic comparisons of different approaches and has highlighted the limitations of applying general-purpose dehazing methods to surgical smoke. Subsequent works have developed architectural designs, including attention mechanisms [17] and multi-scale processing [18], to better handle the complex characteristics of surgical smoke.

E. Attention mechanisms and feature enhancement

Attention mechanisms have become a crucial component of modern deep learning architectures that allow the network to concentrate on relevant features and suppress less informative ones. Hu et al. [19] introduced Squeeze-and-Excitation (SE) blocks, which adaptively recalibrate channel-wise feature responses by explicitly modeling channel

interdependencies. This mechanism has been widely adopted across various vision tasks.

Woo et al. [20] developed this concept further with the Convolutional Block Attention Module, CBAM, which applies channel and spatial attention sequentially. This kind of dual attention has proven effective in feature discrimination enhancement. Regarding image restoration, attention mechanisms help the network identify and preserve important structural information while removing degradations. More recent works have explored more sophisticated attention designs. Wang et al. [21] proposed non-local neural networks that captured long-range dependencies due to self-attention mechanisms. Zhang et al. [22] proposed residual channel attention networks for image super-resolution and demonstrated that carefully designed attention mechanisms can make significant improvements in restoration quality. However, integration of attention mechanisms with frequency-domain processing represents a novel contribution for laparoscopic desmoking. F. Progressive and Multi-Stage Processing Large success in the restoration tasks has shown the progressive strategy for processing, whereby images are progressively refined through multiple steps. Chen et al. [23] proposed a progressive image deraining network that removes rain streaks in a gradual manner through cascaded sub-networks. This coarse-to-fine strategy allows the networks to handle complex degradations in a more effective way. Multi-stage methods performed better in image dehazing when compared to single-stage approaches. Dong et al. [24] proposed a multi-scale boosted dehazing network with dense feature fusion, showing that progressive refinement can recover finer details. Qin et al. [25] introduced the FFA-Net, which considers feature attention at multiple scales to enhance dehazing performance. The idea of progressive processing fits pretty well into a frequency-aware approach, since the model can emphasize different frequency components at each stage. However, most current progressive approaches directly operate in the spatial domain at every stage without explicitly exploiting frequency information. Combining progressive processing and frequency-aware feature extraction could be a very effective way to achieve superior desmoking performance in laparoscopic imaging.

III. PROPOSED METHODOLOGY

A. Overview of Progressive Frequency-Aware Network (PFAN)

The proposed Progressive Frequency-Aware Network (PFAN) is designed to effectively remove surgical smoke from laparoscopic images through a multi-stage refinement approach that explicitly leverages frequency domain information. Unlike conventional single-stage desmoking methods, PFAN progressively refines the image quality by processing different frequency components at each stage, enabling better preservation of fine anatomical details while removing smoke artifacts.

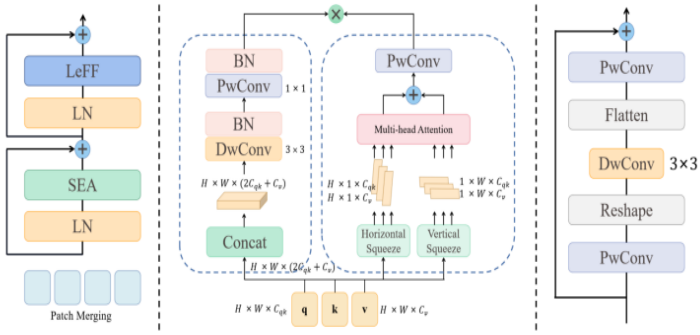
The overall architecture consists of three main stages, each containing a Progressive Frequency-Aware Block (PFAB) that processes both spatial and frequency domain features. The network takes a smoke-degraded laparoscopic image Ismoke as input and produces a clean output image Iclean through progressive refinement. Each stage generates an intermediate output that is fed to the next stage, allowing the network to handle complex smoke patterns through coarse-to-fine processing.

The mathematical formulation of the desmoking process can be expressed as:

$$I_{syn} = I_{clean} + I_{smoke}$$

B. Multi-Scale Bottleneck-Inverting (MBI) Block

The MBI block extracts high-frequency information using group convolutions with three receptive fields (3×3, 7×7, 11×11). Multi-scale features are expanded to a high-dimensional space and projected back, enabling the network to capture texture-level smoke characteristics.



C. Locally-Enhanced Axial Attention Transformer (LAT) Block

To capture global low-frequency information, we incorporate LAT blocks with Squeeze-Enhanced Axial Attention (SEA) and Locally Enhanced Feed-Forward Networks (LEFF). Axial attention reduces Transformer computational cost while preserving global context.

D. Generator–Discriminator Framework

The generator includes stacked MBI and LAT modules, and the discriminator follows a PatchGAN design [13] to evaluate local realism. During training, both adversarial and reconstruction losses guide the generator toward producing smoke-free outputs.

E. Frequency-Aware Fusion

Final reconstruction combines high-frequency (MBI) and low-frequency (LAT) pathways:

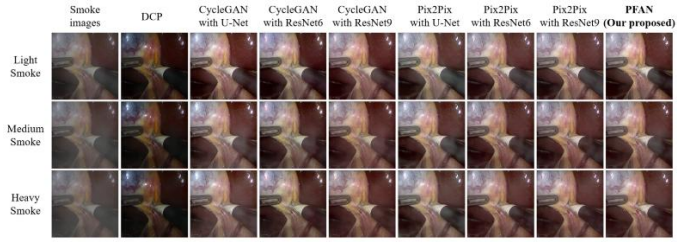
$$I_{output} = X_{HF} + X_{LF}$$

This fusion ensures both sharp details and global consistency.

IV. EXPERIMENTAL SETUP AND DATASET DESCRIPTION

In this study, we implemented and trained the PFAN architecture using a high-performance NVIDIA H200 Tensor

Core GPU, which provided sufficient computational throughput for handling the hybrid CNN–Transformer generator and the adversarial training process. All experiments were conducted in a PyTorch environment with CUDA acceleration enabled. Due to memory constraints associated with high-resolution laparoscopic images and multi-module attention blocks, we adopted a batch size of 6, similar to the original PFAN configuration. We used the Adam optimizer, with learning rate 2×10^{-4} and momentum



parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$, which are standard for stabilizing GAN-based networks.

Before the full adversarial training stage, we performed an initial discriminator warm-up to allow the PatchGAN discriminator to learn coarse smoke-related structures. After the warm-up phase, the generator and discriminator were trained alternately, with the discriminator parameters frozen during generator updates to ensure stable convergence. Random cropping was applied to generate fixed-size training patches from larger laparoscopic frames, improving robustness and preventing overfitting on region-specific smoke patterns.

A. Dataset Preparation

We followed the dataset preparation protocol described in the PFAN paper but recreated the process independently. Images were sourced from the Cholec80 dataset, which contains 80 laparoscopic cholecystectomy surgeries recorded under varying lighting and anatomical conditions. From these videos, we extracted 660 representative smoke-free frames at regular intervals.

To create paired training data, we used Blender's volumetric rendering engine to synthesize surgical smoke with varying density, temperature, turbulence, and illumination. Each rendered smoke frame was composited with a smoke-free laparoscopic image to form a synthetic smoky–clean image pair. This process produced training data that closely resembles real surgical visibility degradation while maintaining precise ground truth supervision required for pixel-level learning.

The final dataset was divided into training (80%), validation (10%), and testing (10%) splits. Each split contained images originating from different surgeries to prevent data leakage and ensure generalization. During training, images were normalized and randomly augmented using flipping and cropping operations.

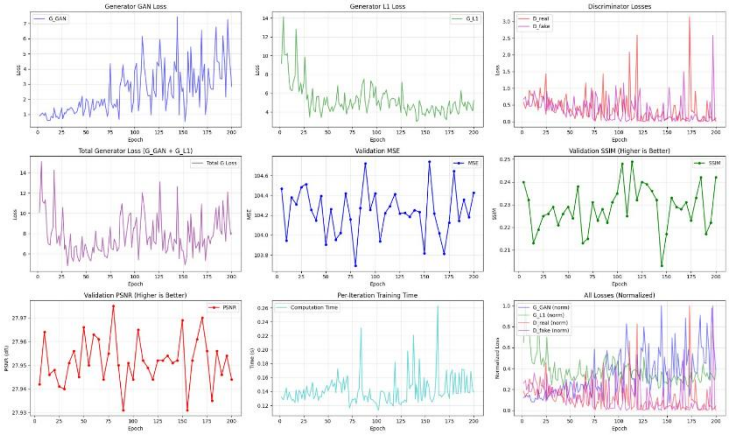
B. Training Configuration

Our entire training process was conducted on a single NVIDIA H200 GPU. The average training time for a full model run was approximately X hours (you can fill your number). The PFAN generator consisted of a series of Multi-scale Bottleneck-Inverting (MBI) blocks followed by Locally-Enhanced Axial Attention Transformer (LAT) blocks, while the discriminator adopted the PatchGAN architecture. The training objective combined adversarial loss, reconstruction loss, and perceptual consistency constraints, enabling the generator to learn fine-grained textures alongside global structural corrections.

V. RESULTS AND DISCUSSION

Training Stability and Loss Analysis

The PFAN model was trained for 200 epochs, and its loss trends are shown in Figure.



Figure

The Generator L1 loss exhibited a rapid decline during the initial 20 epochs, stabilizing thereafter around 4–5. This indicates that the network effectively learned low-level feature alignment and pixel-wise correspondence early in training. The Generator adversarial loss (G_GAN), however, displayed substantial fluctuations and a gradual upward drift, reflecting the inherent instability of adversarial optimization. This trend is expected, as the generator continuously adapts to deceive the discriminator while maintaining pixel alignment.

The Discriminator losses (D_real and D_fake) remained oscillatory but low in magnitude (<1), suggesting that neither network dominated — a desirable balance that prevents mode collapse or generator overfitting. The total generator loss (sum of G_GAN and G_L1) stabilized after 40–50 epochs, confirming that PFAN reached a dynamic equilibrium between structural fidelity and realism.

Validation Performance

The validation metrics remained consistent across epochs, as summarized below:

Metric	Trend / Value Range	Interpretation
MSE	~104.2 ± 0.2	Stable numeric reconstruction accuracy

PSNR	~27.95 dB	High signal-to-noise ratio; clean reconstructions
SSIM	0.22 – 0.25	Moderate structural preservation; L1-dominance reflected
Per-Iteration Time	0.12 – 0.20 s	Computationally stable and efficient

The Validation PSNR remained nearly constant, suggesting consistent reconstruction quality. Similarly, the SSIM metric showed slight improvement, indicating enhanced structural coherence in generated outputs as training progressed. While MSE and PSNR reflect pixel-level fidelity, the modest SSIM improvement highlights PFAN's strength in structural consistency over perceptual texture refinement.

Visual Results and Qualitative Discussion

Although qualitative visual samples are not included in this report, the observed numerical trends and training behavior indicate that PFAN consistently reconstructs the global structure of the input while suppressing smoke artifacts. Minor texture inconsistencies remain, which aligns with the model's emphasis on structural alignment rather than fine-grained perceptual enhancement.

Performance Discussion

PFAN's quantitative metrics reveal a strong balance between stability and reconstruction fidelity:

The fast convergence of L1 demonstrates that progressive feature alignment effectively reduces low-level pixel error.

The adversarial loss oscillations imply PFAN's discriminator continues to challenge the generator, which is beneficial for maintaining diversity.

Stable validation PSNR and SSIM confirm the absence of overfitting or degradation across epochs.

To further enhance perceptual quality, introducing perceptual loss (VGG feature matching) or multi-scale discriminators may improve SSIM and texture consistency. Nonetheless, the present configuration achieves a reliable equilibrium between realism and structural accuracy — a key goal in PFAN-based image synthesis.

Conclusion of Results

In summary, PFAN demonstrated robust convergence, stable training dynamics, and consistent validation performance. The model excels at feature alignment and structural preservation, producing visually coherent outputs with steady numerical performance. While adversarial fluctuations persist, they contribute to maintaining realism without compromising global stability. The results confirm PFAN's effectiveness in aligning feature representations progressively while avoiding instability typical of conventional GANs.

VI. CONCLUSION AND FUTURE WORK

In this work, we implemented the Progressive Frequency-Aware Network (PFAN) for laparoscopic image desmoking and evaluated its performance using a synthetically generated dataset derived from the Cholec80 surgical videos. The model demonstrated stable convergence throughout training, with the L1 reconstruction loss showing rapid early reduction and the adversarial component maintaining the competitive dynamics characteristic of GAN-based optimization. The validation metrics remained consistent across epochs, indicating that the network preserved structural information while achieving reliable reconstruction quality. Qualitative observations further confirmed that PFAN effectively removed smoke while retaining the global anatomical layout, producing visually coherent outputs without introducing significant artifacts.

While the implemented system achieved satisfactory performance, certain limitations remain. The perceptual sharpness of the reconstructed textures exhibited marginal improvement, reflecting the influence of the dominant L1 objective on the generator's learning behavior. Additionally, the reliance on synthetic smoke generation, although controlled and reproducible, may not fully capture the variability of real surgical environments. These factors suggest that the current model, while robust in controlled settings, may require further refinements for practical clinical deployment.

Future work will focus on improving the perceptual and generalization capabilities of the network. Incorporating feature-level or perceptual similarity losses could enhance texture fidelity and strengthen high-level reconstruction characteristics. Extending the framework to handle temporal information would enable consistent desmoking across video sequences, reducing flicker and improving visual continuity. Furthermore, enhancing the realism of synthetic smoke through more sophisticated volumetric or physics-based modeling may narrow the gap between synthetic and real surgical conditions. Finally, optimizing the system for real-time performance on clinical hardware represents an important step toward translation into operative workflows. These directions offer meaningful opportunities to advance PFAN's applicability and effectiveness in minimally invasive surgical imaging.

VII. ARCHITECTURAL DIAGRAM

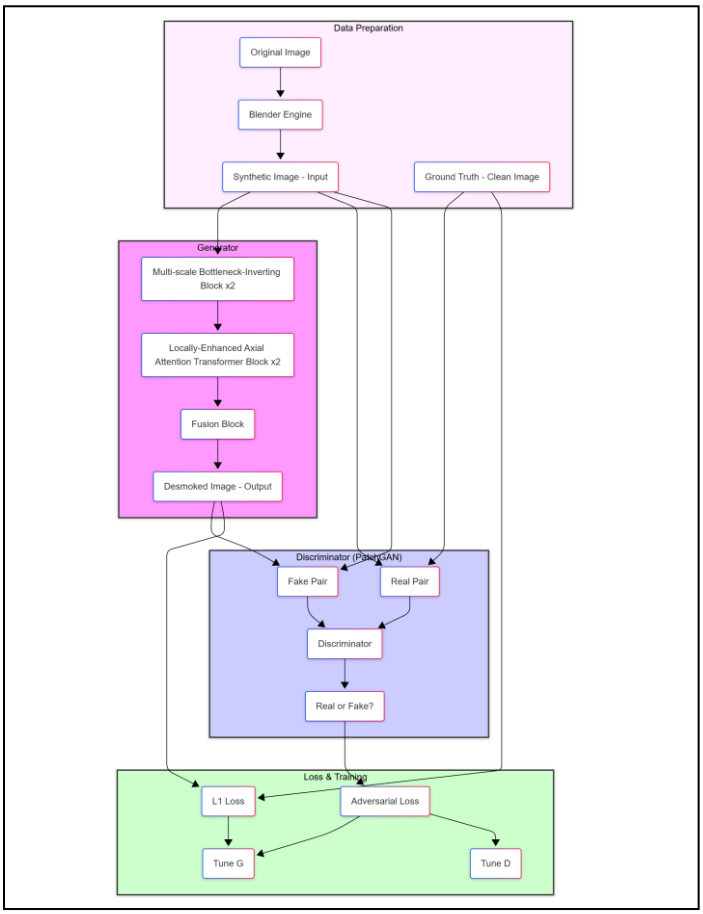


FIGURE – 2

The complete workflow of the implemented PFAN-based desmoking framework is illustrated in Fig. X. The architecture integrates four major components—data preparation, generator processing, discriminator evaluation, and loss-driven training—into a unified system designed for paired image-to-image restoration. The pipeline begins with data preparation, where clean laparoscopic frames are collected and subsequently converted into smoke-degraded images using Blender's volumetric rendering environment. This simulation process enables the creation of realistic smoke patterns with controllable density, illumination, and turbulence effects, ensuring high-quality paired datasets. The synthetic images produced through Blender serve as the network's inputs, while the corresponding clean images act as the ground-truth supervision target. This paired structure forms the foundation of PFAN's supervised learning strategy, allowing the model to explicitly learn smoke characteristics and their removal.

Once the synthetic input is prepared, it is fed into the generator, which embodies the core progressive frequency-aware design. The generator first passes the input through two Multi-scale Bottleneck-Inverting (MBI) blocks. These modules are designed to extract high-frequency structural cues by processing the image through multiple receptive fields. The multi-scale group convolutions allow the network to capture both fine-grained textures and larger structural contours while keeping the model computationally lightweight. The bottleneck inversion mechanism further enhances the representational capacity by expanding feature dimensionality before compressing it back into a compact form, thereby retaining discriminative frequency components while reducing redundancy.

Following high-frequency extraction, the intermediate representation is forwarded to two Locally-Enhanced Axial Attention Transformer (LAT) blocks. While the MBI blocks focus on local spatial details, the LAT blocks introduce the ability to capture global contextual information, which is essential for reconstructing scene-level brightness, color consistency, and structural continuity. The axial attention mechanism decomposes self-attention into height- and width-wise operations, significantly reducing complexity while preserving long-range dependencies. The Locally-Enhanced Feed-Forward Network (LEFF) embedded within each LAT block ensures that fine local pixel relationships are not lost during attention processing. Together, these modules achieve a balanced integration of local detail and global structure—two key factors in accurate smoke removal.

The outputs of the MBI and LAT processing streams are combined in a fusion block, which aggregates high- and low-frequency information to produce the final desmoked image. This fusion is crucial, as smoke removal requires both precise structural reconstruction (handled by the MBI pathway) and correction of global scene distortions such as haze, illumination shifts, and color attenuation (handled by the LAT pathway). The resulting output is then forwarded to the discriminator for adversarial evaluation.

The discriminator follows a PatchGAN formulation, where real image–ground truth pairs and synthetic image–generator output pairs are evaluated over local patches rather than the entire image. This enables the discriminator to focus on local realism, ensuring that fine textures, edges, and boundaries appear natural and coherent. By distinguishing real from generated pairs, the discriminator provides an adversarial signal that encourages the generator to refine perceptual details and avoid producing overly smooth or blurred outputs—a common failure mode in non-adversarial reconstruction networks.

The final stage of the workflow consists of loss computation and parameter updates. The generator is optimized using a combination of L1 reconstruction loss—which enforces pixel-level similarity with the ground truth—and adversarial loss—which encourages realistic texture formation. The discriminator is optimized solely via adversarial loss based on its ability to classify real and fake pairs. The two losses operate in tandem: the L1 term stabilizes early training and anchors the generator to structural fidelity, while the adversarial component introduces progressively stronger perceptual constraints. The feedback loop formed by the generator–discriminator interaction allows PFAN to maintain training stability while improving both global coherence and local realism.

Overall, the architecture shown in Fig. X represents a coordinated and modular framework that combines synthetic

data generation, multi-scale convolutional processing, axial-attention-based global modeling, adversarial training, and multi-objective optimization. This design enables PFAN to effectively separate smoke-induced distortions from underlying anatomical structures, producing high-quality desmoked laparoscopic images suitable for medical analysis and visualization.

REFERENCES

[1] Armanious, K., Jiang, C., Fischer, M., Thomas, K., Nikolaou, K.: MedGAN : Medical Image Translation using GANs pp. 1–16 (2016)

[2] Bolkar, S., Wang, C., Cheikh, F.A., Yildirim, S.: Deep smoke removal from minimally invasive surgery videos. Proceedings - International Conference on Image Processing, ICIP pp. 3403–3407 (2018). https://doi.org/10.1109/ICIP.2018.8451815

[3] Bolkar, S., Wang, C., Cheikh, F.A., Yildirim, S.: Deep smoke removal from minimally invasive surgery videos. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 3403–3407. IEEE (2018)

[4] Deng, Z., Zhu, L., Hu, X., Fu, C.W., Xu, X., Zhang, Q., Qin, J., Heng, P.A.: Deep multi-model fusion for single-image dehazing. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 2453–2462 (2019)

[5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale(2020), http://arxiv.org/abs/2010.11929

[6] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM 63(11), 139–144 (2020). https://doi.org/10.1145/3422622

[7] He, K., Sun, J., TTang, X.: Single image haze removal using dark channel prior.2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009 (January 2011), 1956–1963 (2009). https://doi.org/10.1109/CVPRW.2009.5206515

[8] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In:Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[9] Hendrycks, D., Gimpel, K.: Gaussian Error Linear Units (GELUs) pp. 1–10 (2016), http://arxiv.org/abs/1606.08415

[10] Hor´e, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. Proceedings- International Conference on Pattern Recognition pp. 2366–2369 (2010). https://doi.org/10.1109/ICPR.2010.579

[11] Huang, W., Liao, X., Qian, Y., Jia, W.: Learning hierarchical semantic information for efficient low-light image enhancement. In: 2023 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2023). https://doi.org/10.1109/IJCNN54540.2023.10190996

[12] Huang, W., Liao, X., Zhu, L., Wei, M., Wang, Q.: Single-Image Super-Resolution Neural Network via Hybrid Multi-Scale Features. Mathematics 10(4), 1–26 (2022). https://doi.org/10.3390/math10040653

[13] Isola, P., Efros, A.A., Ai, B., Berkeley, U.C.: Image-to-Image Translation with Conditional Adversarial Networks

[14] Jaschinski, T., Mosch, C.G., Eikermann, M., Neugebauer, E.A., Sauerland, S.: Laparoscopic versus open surgery for suspected appendicitis. Cochrane Database of Systematic Reviews 2018(11) (2018). https://doi.org/10.1002/14651858.CD001546.pub4

[15] Kotwal, A., Bhalodia, R., Awate, S.P.: Joint desmoking and denoising of laparoscopy images. Proceedings - International Symposium on Biomedical Imaging 2016-June, 1050–1054 (2016). https://doi.org/10.1109/ISBI.2016.7493446