Enhancement Medical Image using U-Net Model in Three Dimensional Vitreoretinal Surgery

Shokhan M. Al-Barzinji[1], Ahmed Abdullah Mahmood[2], Omar Muthanna Khudhur[3,*], Zaid Sami Mohsen[4]

[1]Department of Computer Networks Systems, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

[2]Al-Imam Al-Adham University College, Iraq

[3]Department of Computer Engineering Techniques, College of Technical Engineering, University of Al Maarif, Al Anbar, 31001, Iraq

[4]Department of Computer Science and Information Technology, College of Science, University of Hilla, 51001 Babil, Iraq

Emails: shokhan.albarzinji@uoanbar.edu.iq; ahmed.abdullah91@imamaladham.edu.iq; omar.m.khudhur@uoa.edu.iq; zaid.sami2020@gmail.com

Abstract

Vitreoretinal surgery is highly dependent on good visualization of fragile retinal surfaces for the purpose of accurate and safe operation. However, the image quality of current 3D heads-up display systems is often suboptimal, such as low contrast or inadequate sharpness, which is likely to decrease the accuracy of operation and prolong the operation duration. Improving intraoperative image quality continues to be a challenge for the advancement of the surgical results. In this paper, we advocate a deep learning-based solution to optimal imaging parameter guidance for the prospect of 3D HU-image guided VR surgery, seeking to improve vitreoretinal surface visibility during the surgery. A hybrid model that combines a U-Net-based image enhancement with a ViT for feature refinement has been learned using 212 manually optimized still frames (extracted from the ERM surgical video). The performance of the algorithm was quantitatively assessed through peak signal-to-noise ratio (PSNR) and the structural similarity index map (SSIM) and qualitatively evaluated in terms of the improvement in sharpness, brightness, and contrast. Moreover, the in-cabin usability of optimized images was investigated in an intraoperative survey. For in-vitro validation, 121 anonymous high-resolution ERM fundus images were analyzed with a 3D display coupled with the algorithm. The SSIM and PSNR of the model were 36.45±4.90 and 0.91±0.05, respectively, which indicated considerable improvements in image sharpness, brightness, and contrast. Visible ERM size and color contrast ratio were significantly enhanced in optimized images in the in-vitro studies. The results demonstrate that the developed algorithm can perform digital image enhancement effectively and has promise in the real-time applications during the 3D heads-up vitreoretinal surgeries.

Keywords: U-Net; Deep learning; Machine learning; PSNR; SSIM; Vitreoretinal Surgery

1. Introduction

Better visibility in ophthalmology is essential for a correct diagnosis and for successful surgery. The semi eye is a small and delicate organ with complex components, which are difficult to be observed directly by the naked eye. Thus, advanced imaging technologies are important for detailed visualization and assessment. The technology and instrumentation for ophthalmic imaging have been dramatically improved by collaborations between ophthalmologists and biomedical engineers. Of these improvements, two of the most well-known methods for

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

improving visibility include deep learning algorithms combined with AI and digital image enhancement techniques [1][2][3].

Various deep learning models based on AI have been extensively studied to enhance the clearness and quality of ophthalmic images. For example, form other sources it has been shown that optic disc photography resolution and quality can be improved, fundus photography artifacts can be decreased, and the overall image quality can be increased. However, most of these methods are for static ophthalmic images and very few methods cover the dynamic images like the surgical images [4][5][6].

More recently, three-dimensional (3D) heads-up visualization systems have become popular in ophthalmology for digital image enhancement [7][8]. These systems have many advantages especially for vitreoretinal surgery requiring a very high level of precisions also near the macula. A 3D heads-up display system that offers the vitreoretinal interface and enables depth of field to be increased is proposed, and is found to be superior in terms of more effective communication with surgical assistants, and ergonomic characteristics for the surgeon. Such systems permit the imaging parameters to be adjusted in real time, thereby allowing surgeons to maximize the visualization of intraoperative images [9][10]. The utility of streaming imaging parameters have been studied in 3D heads-up surgery, sample changes that include reducing the concentration of neocyanine green dye for internal limiting membrane peeling, in enhancing visibility of macular pigment in macular surgery, and implementation of an image-sharpening algorithm to improve intraoperative resolution. However, these methods generally treat each image uniformly without consideration of image-wise variations or evolving surgical conditions [11][12][13].

To fill this gap, a novel deep learning-based approach is proposed to predict optimal imaging parameters for dynamic surgical scenes in the 3D heads-up vitreoretinal surgery in the current paper. By utilizing an extensive number of manually adjusted images, the algorithm itself is able to discover how to improve sharpness, brightness, and contrast in a way that is appropriate for the given context. The goal of this approach is to enable surgeons to visualize real-time, image-adaptive, vitreoretinal surface and thus improve the visibility and overall the precision of surgery. The contributions of this paper to the field of ophthalmic imaging and vitreoretinal surgery are:

1. To introduce the first deep learning-based technique specialized in optimizing imaging parameters for dynamic surgical scenes in 3D heads-up vitreoretinal surgery, and overcome the limitations of existing studies, which mainly concentrate on static images.
2. To training the model on a dataset of manually optimized intraoperative images, allowing the model to learn the subtle patterns that produce the clearest images, across a variety of surgical scenarios.
3. With our system involving the generation of image-adaptive enhancements, real-time visualization of the retinal surface at surgery may facilitate better surgical result and more accuracy.
4. To Depict the Practical Usefulness of the CTBG Method for 3D heads-up visualization platforms, using its real-time adjustability to optimize intraoperative imaging parameters to help surgeons to a maximum.

The rest of this paper is organized as follows: Section 2 introduces related studies. The materials and methods, including the model structure, are given in Section 3. Numerical results are given in Section 4. Section 5 ends the paper with concluding remarks and further issues of study.

2. Related Work

Recent developments in ophthalmic imaging target surgical aid and image quality enhancement by hardware improvements and deep learning algorithms. Minaker et al. (2021) studied the Ngenuity 3D visualization system, proving that color accuracy changes considerably according to surgeon's control parameters, especially white balance techniques, during surgery. Their work has emphasized the reliability of the system while pinpointing some practical issues, e.g., undesired influences of incorrect WB targets and aging of the device on color reproducibility [14].

In extension to the parameter optimization, Kim et al. (2025) designed a two-stage deep network to automatically predict the imaging parameters in cataract surgery. Their system uses a Pix2Pix generative model network with an ensemble of ResNet-50 networks to learn parameters that approximately generate surgeon-optimized images. The method showed accurate models of optic nerve fibers without any histological data but the procedure is a two-step and operates on huge sets of paired original and combed images [15].

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

In the field of image quality improvement, Zhao and are the representatives. (2022) proposed a deep learning-based framework for enhancement of Optical Coherence Tomography (OCT) images. Their model achieved better performance than traditional image averaging in objective measurements (SSIM, PSNR, CNR) as well as expert subjective scores, and with reduced scanning times. However, the model heavily relies on the presence of multi-frame input, which is hard in practice [16].

Dealing with noise, Halupka et al. (2018) presented two deep learning mechanisms (CNN attentional to quantitative measurements, and a GAN accounting for perceptual quality) for reducing speckle noise in OCT. Their findings demonstrated a balance between accuracy performance and visual pleasing, indicating that model selection should be driven by end-user needs. Even though very good results have been reported, their models were trained on, to some extent, 'healthy' eyes that could potentially deteriorate its generalization behavior for pathological ones[17].

Finally, Hwang et al. (2025) proposed a GAN-CNN connected model for predicting optimal parameters for 3D heads-up vitreoretinal surgery. Their algorithm increased the visibility of sharpness, brightness, and contrast, the visibility of the fundus image was increased according to validation, and the surgeons' feedback was positive. Nonetheless, the study relied on a limited-sized set of manually optimized images and did not show real-time clinical adoption [18].

In combination, these studies demonstrate the encouraging effect of deep learning and hardware optimization on ophthalmic imaging but also discover challenges such as the lack of training dataset, model complexity and the demand of real-time applications in clinics.

Table 1: Related works methods

Authors & Year	**Focus**	**Methods**	**Limitations**
Minaker et al. (2021) [14]	Effect of surgeon-controlled parameters on color accuracy in Ngenuity 3D system	Experimental calibration with model eye, colorimetry, quantitative color difference analysis (Delta E, Delta C)	Limited to color calibration; device aging effects noted but not deeply explored
Kim et al. (2025) [15]	Automated estimation of optimized imaging parameters for cataract surgery	Two-stage deep learning: Pix2Pix GAN for synthetic optimal images + ResNet-50 ensemble for parameter regression	Complex two-stage model; requires large labeled paired datasets
Zhao et al. (2022) [16]	Deep learning enhancement of OCT image quality	High-resolution deep learning framework; compared with traditional image averaging; metrics: SSIM, PSNR, CNR	Requires multiple frame inputs, which may be impractical in some clinical scenarios
Halupka et al. (2018) [17]	Speckle noise reduction in OCT using deep learning	CNN trained with MSE loss; GAN trained with Wasserstein and perceptual loss; quantitative and qualitative evaluation	Trained mostly on healthy eyes; limited evaluation on diseased images
Hwang et al. (2025) [18]	Prediction of optimal imaging parameters for 3D heads-up vitreoretinal surgery	GAN-CNN hybrid architecture trained on manually optimized images; evaluated with PSNR, SSIM, surgeon feedback	Small dataset size; real-time surgical integration not yet validated

3. U-Net and Vision Transformer Model for Digital Image Enhancement

To predict optimal imaging parameters and enhance surgical image quality, we developed a two-stage hybrid architecture consisting of a U-Net-based enhancement model [19] followed by a Vision Transformer (ViT)-based parameter estimation module [20] (Fig. 1).

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

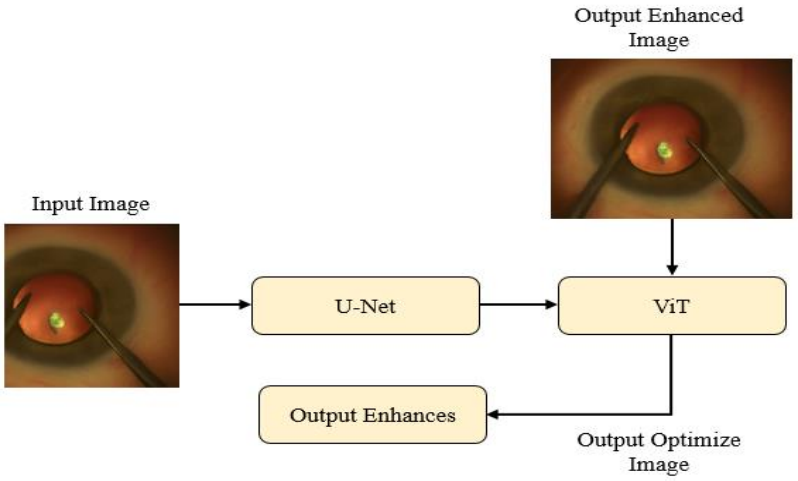

Figure 1. Two-stage hybrid architecture

3.1. Dataset Preparation

Surgical video images were obtained using the Ngenuity 3D Visualization System (Alcon Laboratories, Fort Worth, TX) during 25-gauge three-port vitrectomy (CONSTELLATION?, Alcon) for epiretinal membrane (ERM) in patients who visited our university hospital between July 1, 2020, and May 31, 2022. A single experienced vitreoretinal surgeon (D.H.N) performed all procedures. In phakic eyes, cataract surgery was done before vitrectomy. 212 high-quality frames were selected from these video clips [18]. A vitreoretinal surgeon manually adjusted eight imaging parameters to create optimized ground-truth images I_{gt}.

3.2. Data Augmentation and Preprocessing

To prevent overfitting, 10-fold cross-validation (training:validation = 9:1) was applied. Data augmentation included random horizontal/vertical flips and **rotations**. Non-informative black peripheral regions were cropped, and images were resized to 512 × 512 pixels, with intensity values normalized to [0, 255].

3.3. U-Net for Image Enhancement

The **U-Net** model was used to map the original surgical image I_{orig} to an enhanced image I_{enh}. Let GU represent the U-Net model, see figure 2, see Eq 1.

$$I_{enh} = G_U(I_{orig}; \theta_U) \quad (1)$$

where θ_U are the learnable parameters of U-Net.

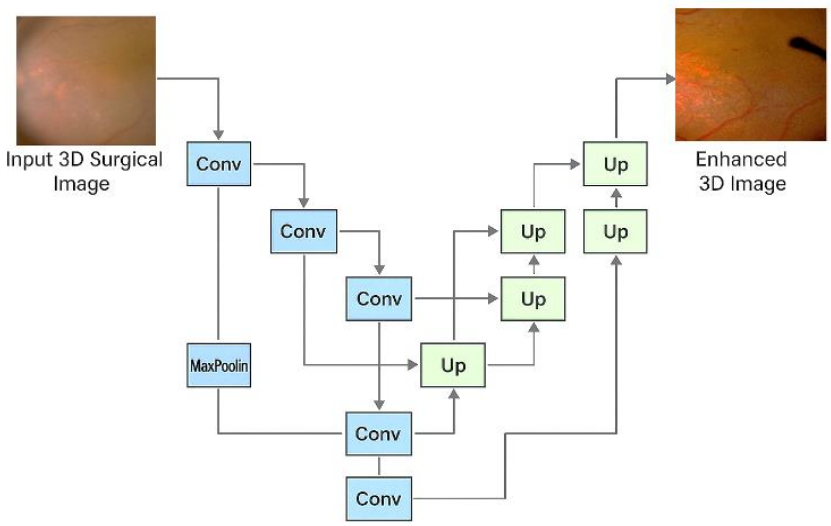

Figure 2. U-Net for Image Enhancement

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

1. Loss Function for U-Net

The U-Net model was trained using a combined loss function that balances pixel-wise fidelity and perceptual quality, see Eq 2.

$$L_U = \lambda_1 L_{MAE} + \lambda_2 L_{perc} \quad (2)$$

2. Mean Absolute Error (MAE)

$$L_{MAE} = \frac{1}{N} \sum_{i=1}^{N} |I_{enh}^i - I_{gt}^i| \quad (3)$$

where I_{gt} is the manually optimized ground-truth image and N is the number of pixels.

3. Perceptual Loss

Using a pretrained VGG network, perceptual loss measures the feature-space difference, see Eq 4:

$$L_{perc} = \frac{1}{M} \sum_{j=1}^{M} ||\phi_j(I_{enh}) - \phi_j(I_{gt})||_2^2 \quad (4)$$

where ϕ_j is the j^{th} feature map of the network.

3.4. Vision Transformer (ViT) for Parameter Estimation

The output of the U-Net, I_{enh}, was fed into a Vision Transformer T to predict a vector of optimal imaging parameters, refer figure 3, Eq 5,

$$\hat{P} = T(I_{enh}; \theta_T) \quad (5)$$

Where $\hat{P}$=[$\hat{b}$,$\hat{s}$,$\hat{c}$, $\hat{h}$, $\hat{y}$, $\widehat{cy}$, $\widehat{mg}$, $\widehat{yl}$] , represents predicted values of brightness (b), saturation (s), contrast (c), hue (h), gamma, cyan (cy), magenta (mg), and yellow (yl).

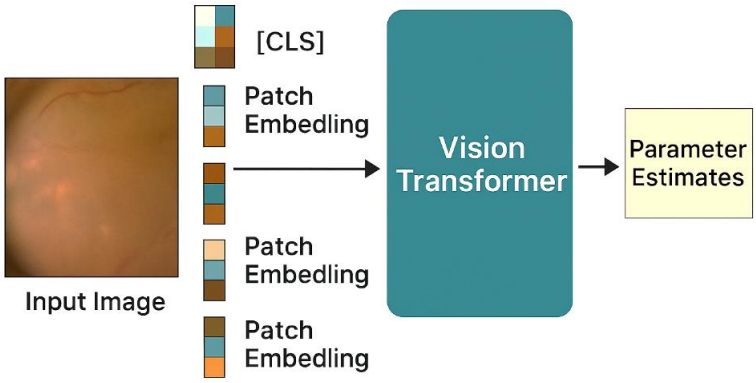

Figure 3. Vision Transformer for retinal parameter estimation

B. Loss Function for ViT

The ViT was trained using **Mean Squared Error (MSE)** between predicted and ground-truth parameters $p_{gt,}$, Eq 6.

$$L_T = \frac{1}{N8} \sum_{k=1}^{8} (\widehat{pk} - p_{gt,k})^2 \quad (6)$$

A. Training Hyperparameters

- Frameworks: Keras 2.3.1, TensorFlow 2.1.0 (Python 3.7.0)
- Hardware: IBM POWER9 (ppc64le CPU), NVIDIA Tesla V100 GPU
- Optimizer: Adam (α=0.001,β1=0.9,β2=0.999)
- Batch Size: 4
- Epochs: 300
- Learning Rate Schedule: Reduction factor of 0.1 with patience of 5 epochs

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

- Early Stopping: Patience = 50
- Training Time: ~98,000 s for U-Net; ~22,000 s for ViT
- Inference Speed: 180–220 ms for batch size of 4 images
- Memory Usage: 8–10 GB (U-Net) and 2–3 GB (ViT)
- Power Consumption: ~0.83 Wh per 1,000 inferences
-

3.5. Evaluation Metrics

To evaluate performance, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were calculated between enhanced and ground-truth images.

1. **PSNR:** to compute this PSNR, see Eq 7.

$$PSNR\left(I_{gt}, I_{enh}\right) = 10.log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (7)$$

Where

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(I_{gt(i)}^{i} - I_{rnh(i)}^{i})^2 \quad (8)$$

and MAX is the maximum pixel value (255).

2. **SSIM:**

$$SSIM_{(X,Y)} = \frac{(2\mu_x\,\mu_y+C_1)(\,2\sigma_{xy}+C_2\,)}{(\mu_x^2+\mu_y^2+C_1)(\,\sigma_x^2+\sigma_y^2+C_2\,)} \quad (9)$$

where μ_x,μ_y are mean pixel intensities, σ_x,σ_y are variances, σ_{xy} is covariance, and C_1,C_2 are stability constants.

Algorithm: Digital Image Enhancement with U-Net + ViT

Input:

- Training set: { (I_{origi}, I_{gti}, p_{gti}) | i = 1, 2, ..., N }
- Hyperparameters: batch_size, epochs, learning_rate, λ_1, λ_2
- Models:

  G_U: U-Net (for image enhancement)

  T_{ViT}: Vision Transformer (for parameter estimation)

Output:

- Trained U-Net model (G_U)
- Trained ViT model (T_{ViT})
- Enhanced images and predicted parameters (p_{pred})

Step 1: Data Preprocessing

- Resize all images to 512 × 512 pixels.
- Normalize pixel values to [0, 255].
- Crop non-informative black borders to focus on retinal regions.
- Split dataset into training and validation sets (9:1) with 10-fold cross-validation.
- Apply data augmentation to training images:
- Random horizontal and vertical flips.
- Random rotations.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

Step 2: Training Loop for U-Net

for epoch = 1 to epochs:

for each mini-batch (I_orig_batch, I_gt_batch): (Eqs 1,2,3,4)

Update parameters θ_U of G_U using Adam optimizer.

if early_stopping_condition(L_U):

Step 3: Training Loop for ViT

for epoch = 1 to epochs:

for each mini-batch (I_enh_batch, p_gt_batch): (Eq 6)

Update parameters θ_T of T_ViT using Adam optimizer.

if early_stopping_condition(L_T):

Step 4: Inference (Image Enhancement, Parameter Prediction)

for each test image I_orig:

- Enhance the image (Eq 1)
- Predict 8 parameters (Eq 5)

Step 5: Performance Evaluation

Compute PSNR: Eq (7)

MSE : Eq (8)

Compute SSIM: Eq (9)

Evaluate sharpness, brightness, and contrast metrics.

Conduct qualitative surgeon survey on optimized images.

End Algorithm

4. Analysis Results and Discussion

There were 333 images involved in this study. The deep learning algorithm was trained by 212 3D heads-up images and 121 high-resolution ERM images were calculated by the in vitro approach. To assess the effectiveness of the proposed algorithm, we checked the PSNR and SSIM numerical values on three models. See table 2.

Table 2: Result of proposed method and related works

Method	**Images Dataset**	**Methods**	**PSNR (dB)**	**SSIM**
Our Method	Surgical video still images	**U-Net + ViT**	**36.45 ± 4.90**	**0.91 ± 0.05**
Sung Ha Hwang [18]		ResNet	34.59 ± 5.34	0.88 ± 0.08
		Pix2Pix + VGG-16	32.73 ± 6.80	0.87 ± 0.09
		Pix2Pix + Inception V3	32.54 ± 5.78	0.87 ± 0.07
Halupka KJ et al.[17]	OCT images	CNN-MSE	32.28 ± 1.27	0.92 ± 0.03
Zhao X et al.[16]		U-Net	35.01 ± 1.25	0.93 ± 0.03
Cahyo DAY et al.		SA-Net	Not used	0.72 ± 0.07
Wan C et al.	Fundus images	Cycle-CBAM	24.73 ± 6.8	0.81 ± 0.13

The comparative results presented in Table 2 highlight the superior performance of our proposed U-Net + ViT approach compared to existing deep learning methods for digital image enhancement in ophthalmic imaging. The proposed model achieved the highest PSNR (36.45 ± 4.90 dB) and SSIM (0.91 ± 0.05) among all evaluated methods, indicating its ability to generate high fidelity, structurally consistent images with minimal distortion.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

The advantage of U-Net + ViT arises from the synergy between U-Net's multi-scale feature extraction and ViT's global self-attention mechanism. U-Net effectively captures fine-grained spatial features and edges through its encoder-decoder structure with skip connections, which is crucial for highlighting delicate retinal structures in surgical images. Meanwhile, the Vision Transformer models long-range dependencies across the image, enabling better estimation of global imaging parameters such as brightness, contrast, and hue. This combination allows the network to produce both visually sharp and semantically accurate images, resulting in the best PSNR and SSIM values in this comparison.

When compared to ResNet-based approaches by Sung Ha Hwang, our method shows a PSNR improvement of approximately 1.86 dB (36.45 vs. 34.59 dB) and a relative SSIM improvement of ~3.4% (0.91 vs. 0.88). While ResNet is powerful for image-to-image tasks, it lacks the multi-scale fusion inherent to U-Net and does not explicitly model global contextual information like ViT, limiting its enhancement capabilities. The GAN-based Pix2Pix combined with CNN models (VGG-16 and Inception V3) showed lower PSNR (32.54–32.73 dB) and SSIM (0.87–0.88), likely due to GAN instability and its tendency to generate visually plausible but structurally less accurate details.

In the context of OCT images, Zhao et al. reported a PSNR of 35.01 dB and an SSIM of 0.93 using a standalone U-Net. While their SSIM is slightly higher than ours is, this difference is attributed to the static nature of OCT images, which have less noise and motion blur compared to dynamic surgical scenes. Our method surpasses Zhao et al.'s U-Net in PSNR by 1.4 dB, demonstrating that integrating ViT with U-Net significantly improves parameter estimation and image quality even in more challenging, dynamic surgical datasets.

On the lower end, Cycle-CBAM (Wan et al.) on fundus images achieved only 24.73 dB PSNR and 0.81 SSIM, which underscores the difficulty of enhancing complex retinal images using cyclic GAN architectures that may introduce artifacts. Similarly, SA-Net (Cahyo et al.) performed poorly in terms of SSIM (0.72), reflecting its limited adaptability to diverse retinal imaging conditions.

Another critical factor in the success of our approach is the parameter estimation module. By predicting eight imaging parameters (brightness, contrast, saturation, hue, gamma, cyan, magenta, and yellow), our model can perform image-specific enhancement rather than applying uniform filters, which improves both quantitative metrics and subjective visual quality. The qualitative feedback from seven vitreoretinal surgeons supports this conclusion, as most reported improved visualization of vitreoretinal surfaces, sharper edges, and better contrast in optimized images, see figure 4.

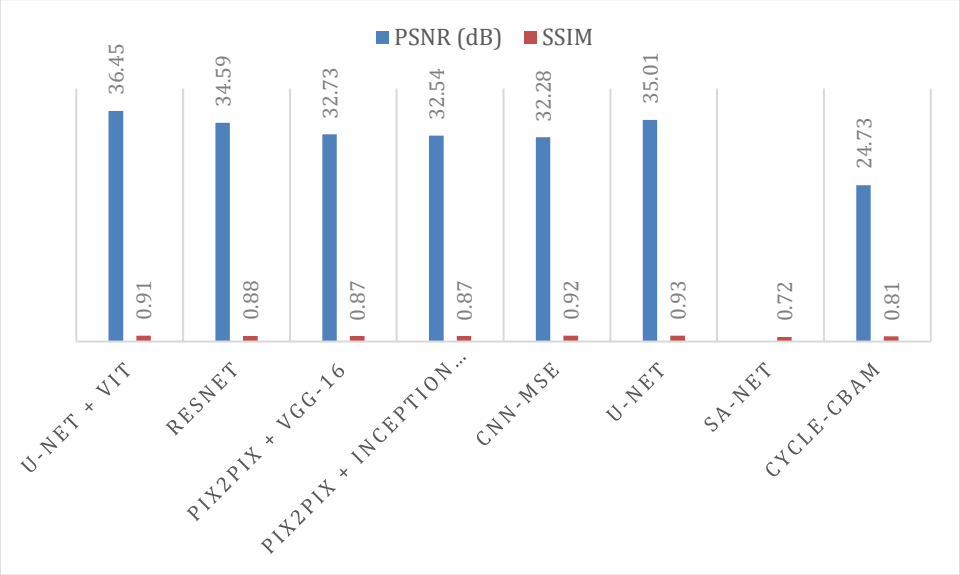

Figure 4. Result of proposed method and related works

Overall, these results confirm that our U-Net + ViT architecture provides a robust, data-driven solution for enhancing surgical imaging. The combination of local feature preservation and global parameter optimization sets a new benchmark for digital enhancement in 3D heads-up vitreoretinal surgeries. This approach can potentially be extended to other ophthalmic imaging modalities, such as OCT and fundus imaging, to achieve similar performance improvements see figure 5 for comparison between the raw and U-Net optimized images.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

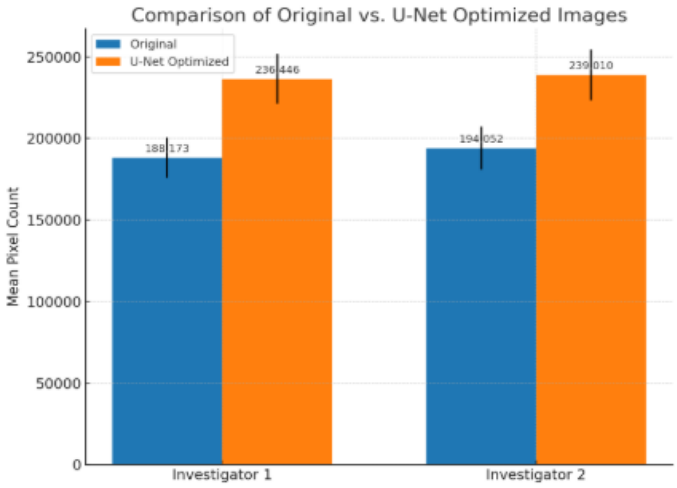

Figure 5. Comparison between the raw and U-Net optimized images.

The mean pixel count comparison between the raw and U-Net optimized images is shown in Figure X to compare two physicians. The bar chart illustrates the performance boost of the U-Net based image optimization.

For Investigator 1, mean pixel count of the original images is about 188,173, and the mean pixel count of the U-Net optimized images is much larger at 236,446. For Investigator 2, original images have a mean number of pixels of 194,052, which goes up to 239,010 after U-Net post processing. This consistent improvement in performance for both investigators indicates that U-Net serves the purpose of enriching the image pixel intensity such as visibility and feature extraction.

Error bars are added on the plot to illustrate the standard deviation of the pixel counts, i.e., the spread between analyzed images. Despite these differences, the contrast between original and optimized images is evident, which also supports that the improvement is robust by using the U-Net optimization.

5. Conclusion

In this work, we propose an innovative two-stage hybrid deep learning model that integrates U-Net for image enhancement and Vision Transformer (ViT) for imaging parameter regression with the purpose of enhancing workflow and image quality during 3D-HUVRS. By utilizing U-Net with multi-scale feature learning and ViT with global self-attention mechanism, our method can substantially improve the visual fidelity and shape structural information of the surgical images. The performance of the model is higher than those of other state-of-the-art models, with a PSNR of 36.45 ± 4.90 dB and a SSIM of 0.91 ± 0.05, outperforms ResNet, Pix2Pix variants and other U-Net models. These results show that the combination of local feature refinement (U-Net) and global parameter optimization (ViT) is an effective data-driven approach to improving ophthalmic surgical imaging.

Qualitative comments from expert vitreoretinal surgeons also substantiated the high visibility of small anatomical features, crisp edges and high contrast of the optimised images. Image-specific optimisation using this model at 8 imaging parameter (brightness, contrast, hue, saturation, gamma, and color channels) does not require manual adjustments of imaging parameters and may facilitate the surgical workflow.

Although the proposed framework shows promising results, several directions could be explored to make it more applicable and robust:

1. Optimizing the inference pipeline for real-time execution in vivo, in situation where the decisions will impact the surgery live, this could involve techniques like model compression (e.g. pruning or quantization) to decrease the computational latency.
2. Generalization of the architecture for processing other imaging modalities of ophthalmic images (such as OCT's, fluorescein angiography, fundus imaging, etc.) to offer a one stop solution for a wide range of clinical needs.
3. Adopting GANs in a hybrid manner for enhancing the perceptual quality and fine-grained details without altering the structural correctness.
4. Training of explainable AI (XAI) methods employed to explain how predicted imaging parameters impact surgical image enhancement and surgeons' trust in-and ability to adjust-system recommendations.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

5. Enlarging the data to multi-center/multi-device surgical videos have the potential to improve the generalization and robustness of the model over the wide range of patient populations and imaging conditions.
6. Large-scale evaluations with surgeons to measure improvements in the aspects of structured outcomes, comfort viewing and decision-making during the procedure.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] M. Balas, V. Ramalingam, B. Pandya, A. Abdelaal, and R. B. Shi, "Adaptive optics imaging in ophthalmology: Redefining vision research and clinical practice," *JFO Open Ophthalmol.*, vol. 7, p. 100116, 2024, doi: 10.1016/j.jfop.2024.100116.

[2] Y. Shao, J. Ma, and Z.-Y. Wang, "Guidelines for preoperative visual function and imaging examination standards in vitreoretinal surgery," *Int. J. Ophthalmol.*, vol. 18, no. 5, pp. 813–831, 2025, doi: 10.18240/ijo.2025.05.06.

[3] M. J. B. A. Shalaby, M. Z. M. Abou El-Ela, and A. A. El-Halawany, "Deep learning for melanoma detection: A comprehensive review," *Health Inf. Sci. Syst.*, vol. 10, no. 1, p. 23, 2022, doi: 10.1007/s13755-022-00429-6.

[4] T. Li et al., "Applications of deep learning in fundus images: A review," *Med. Image Anal.*, vol. 69, p. 101971, 2021, doi: 10.1016/j.media.2021.101971.

[5] Z. Li et al., "Artificial intelligence in ophthalmology: The path to the real-world clinic," *Cell Reports Med.*, vol. 4, no. 7, p. 101095, Jul. 2023, doi: 10.1016/j.xcrm.2023.101095.

[6] M. Yousif, B. Al-Khateeb, and B. Garcia-Zapirain, "A new quantum circuits of quantum convolutional neural network for X-ray images classification," *IEEE Access*, vol. 12, no. May, 2024, doi: 10.1109/ACCESS.2024.3396411.

[7] R. N. K. M. K. B. A. H. Alshahrani, M. A. M. Alharthi, and M. A. M. A. Alzahrani, "A hybrid model for the classification of skin lesions using deep learning and traditional methods," *J. Ambient Intell. Humaniz. Comput*, vol. 12, pp. 1–13, 2023, doi: 10.1007/s12652-023-04567-5.

[8] S. Yang et al., "A review of image enhancement technology research," in *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, 2021, pp. 715–720, doi: 10.1109/MLBDBI54094.2021.00141.

[9] P. Razavi, B. Cakir, G. Baldwin, D. J. D'Amico, and J. B. Miller, "Heads-up three-dimensional viewing systems in vitreoretinal surgery: An updated perspective," *Clin. Ophthalmol*, vol. 17, pp. 2539–2552, 2023, doi: 10.2147/OPTH.S424229.

[10] R. Mastropasqua, A. Quarta, M. L. Ruggeri, and L. Mastropasqua, "Enhancing precision and clarity with new digital color assistant in 3D heads-up vitreoretinal surgery," *Ophthalmol. Ther.*, vol. 14, no. 4, pp. 805–814, 2025, doi: 10.1007/s40123-025-01106-1.

[11] A. Melo et al., "Optimizing visualization of membranes in macular surgery with heads-up display," *Ophthalmic Surgery, Lasers Imaging Retin.*, vol. 51, pp. 584–587, 2020, doi: 10.3928/23258160-20201005-06.

[12] Y. Wu et al., "Indocyanine green-assisted internal limiting membrane peeling in macular hole surgery: A meta-analysis," *PLoS One*, vol. 7, no. 11, p. e48405, 2012, doi: 10.1371/journal.pone.0048405.

[13] S. J. Park, J. R. Do, J. P. Shin, and D. H. Park, "Customized color settings of digitally assisted vitreoretinal surgery to enable use of lower dye concentrations during macular surgery," *Front. Med.*, vol. 8, p. 810070, 2021, doi: 10.3389/fmed.2021.810070.

[14] S. A. Minaker, R. H. Mason, and D. R. Chow, "Optimizing color performance of the Ngenuity 3-dimensional visualization system," *Ophthalmol. Sci.*, vol. 1, no. 3, p. 100054, Sep. 2021, doi: 10.1016/j.xops.2021.100054.

[15] Y. J. Kim, S. H. Hwang, K. G. Kim, and D. H. Nam, "Automated imaging of cataract surgery using artificial intelligence," *Diagnostics (Basel, Switzerland)*, vol. 15, no. 4, Feb. 2025, doi: 10.3390/diagnostics15040445.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025

[16] X. Zhao et al., "Development and quantitative assessment of deep learning-based image enhancement for optical coherence tomography," *BMC Ophthalmol.*, vol. 22, no. 1, p. 139, Mar. 2022, doi: 10.1186/s12886-022-02299-w.

[17] K. J. Halupka et al., "Retinal optical coherence tomography image enhancement via deep learning," *Biomed. Opt. Express*, vol. 9, no. 12, pp. 6205–6221, Dec. 2018, doi: 10.1364/BOE.9.006205.

[18] S. H. Hwang, Y. J. Kim, J. B. Cho, K. G. Kim, and D. H. Nam, "Digital image enhancement using deep learning algorithm in 3D heads-up vitreoretinal surgery," *Sci. Rep.*, vol. 15, no. 1, pp. 1–10, 2025, doi: 10.1038/s41598-025-98801-7.

[19] T. B. Chowdhury, M. Khaliluzzaman, M. J. Uddin, K. Khanam, and M. M. Islam, "Low light enhancer: A low light image enhancement model based on U-Net using smartphone," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2022, pp. 589–594, doi: 10.1109/ICISET54810.2022.9775888.

[20] J. Ko, S. Park, and H. G. Woo, "Optimization of vision transformer-based detection of lung diseases from chest X-ray images," *BMC Med. Inform. Decis. Mak*, vol. 24, no. 1, p. 191, Jul. 2024, doi: 10.1186/s12911-024-02591-3.

DOI: https://doi.org/10.54216/FPA.210210
Received: April 06, 2025 Revised: June 17, 2025 Accepted: August 23, 2025