INTERNATIONAL JOURNAL OF ELECTRONICS AND COMPUTER APPLICATIONS

ORIGINAL ARTICLE

OPEN ACCESS
Received: 28-07-2025
Accepted: 30-11-2025
Published: 23-12-2025

Citation: Chaurasiya A, Jadhav H, Naik S, Rijhwani N, Chaudhari A. Detection and Localization of Retinal Pathologies using Jointly Optimized UVCGAN and CNN Classifier on Color Fundus Photographs. 2025; 2(2):86-92. https://doi.org/10.70968/ijeaca.v2i2.ML104

* **Corresponding author.**
d2022.abhishek.chaurasiya@ves.ac.in

Funding: None

Competing Interests: None

ISSN
Electronic: 3048-8257

Detection and Localization of Retinal Pathologies using Jointly Optimized UVCGAN and CNN Classifier on Color Fundus Photographs

Abhishek Chaurasiya[1*], Harsh Jadhav[1], Shreyas Naik[1], Nidhi Rijhwani[1], Abhishek Chaudhari[1]

1 Department of Information Technology, Vivekanand Education Society's Institute of Technology, Mumbai, India.

Abstract

Automated detection and localization of retinal pathologies from color fundus photographs is critical for early diagnosis of vision-threatening diseases. Existing supervised seg- mentation methods require expensive pixel-level annotations, while classification-only approaches lack spatial interpretability. This paper presents an enhanced framework for retinal disease detection and lesion localization by jointly optimizing a UVCGAN (UNet Vision Transformer Cycle-consistent GAN) generator with an EfficientNet-B0 classifier using only image-level labels. The key contribution is the replacement of the conventional CycleGAN generator with a UVCGAN generator that incorporates a Vision Transformer bottleneck within the U-Net encoder- decoder, enabling the model to capture global retinal structure through self-attention mechanisms. A PatchNCE contrastive loss enforces patch-level structural correspondence, improving upon CycleGAN's cycle consistency approach. The generator translates diseased fundus images into healthy counterparts, and the resulting difference maps highlight lesion areas without pixel- level supervision. The EfficientNet-B0 classifier, jointly optimized with the generator, classifies difference maps while its gradient feedback simultaneously improves the generator's lesion removal capability. Experimental evaluation on the ODIR-5K dataset with 832 images for pathologic myopia detection achieved 93.03% accuracy, 96.07% precision, 86.43% recall, 90.99% F1-score, and AUC of 0.984, demonstrating competitive performance on a realistically imbalanced dataset without pretrained classifier weights.

Keywords: Retinal disease detection, UVCGAN, Vision Transformer, Generative adversarial network, Joint optimization, Lesion localization, Fundus photography, Pathologic myopia

Introduction

Retinal diseases such as pathologic myopia (PM), glaucoma, and age-related macular degeneration (AMD) are leading causes of irreversible vision loss, collectively affecting over 300 million people worldwide[(1)]. Color fundus photography provides a non-invasive, cost-effective screening modality, yet manual interpretation by ophthalmologists is time-consuming, subjective, and limited by workforce shortages, particularly in developing nations.

Deep learning has achieved remarkable success in medical image classification[(6)], but two fundamental challenges persist. First, supervised segmentation methods require pixel-level annotations that are prohibitively expensive to obtain from specialist clinicians. Second, classification-only approaches provide disease labels without spatial evidence of where the pathology exists, limiting clinical trust and adoption.

Zhang *et al.*[1] proposed an elegant solution by jointly optimizing a CycleGAN[4] with a CNN classifier for retinal disease detection and localization. The CycleGAN translates diseased images into healthy counterparts, and the difference maps naturally highlight lesion regions using only image-level labels. However, the CycleGAN generator relies exclusively on convolutional operations with limited receptive fields, restricting it to local feature processing and missing the global spatial relationships between the optic disc, macula, and vascular patterns that are diagnostically critical in fundus images.

In this paper, we address these limitations through two key contributions:

1. **UVCGAN Generator:** We replace the CycleGAN generator with a UVCGAN[2] generator that incorporates a Vision Transformer (ViT)[5] bottleneck within the U-Net architecture. The ViT processes the compressed bottleneck features through multi-head self-attention, enabling each spatial location to attend to the entire fundus image, capturing global retinal structure while preserving local details through skip connections.
2. **PatchNCE Contrastive Loss:** Following Park *et al.*[3], we incorporate a patchwise contrastive loss alongside cycle consistency, enforcing that corresponding patches in input and output share similar features at multiple encoder scales. This provides more principled structural preservation than pixel-level cycle consistency alone.

We additionally upgrade the CNN classifier from ResNet-50[6] to EfficientNet-B0[7], achieving better classification accuracy with fewer parameters.

Related Work

A. GAN-based Disease Detection and Localization

Generative adversarial networks[8] have been increasingly applied to medical image analysis. Siddiquee *et al.*[9] proposed Fixed-Point GAN for disease detection by translating diseased images to the healthy domain and using the maximum pixel difference as a detection score. However, without a jointly optimized classifier, performance is vulnerable to outliers in the difference maps.

Zhang *et al.*[1] introduced a joint optimization framework combining CycleGAN with a CNN classifier, where the classifier's cross-entropy loss backpropagates through the generator, creating a cooperative feedback loop. Their res-guided U-Net generator with pixel-adaptive convolutions achieved state-of- the-art results on LAG (glaucoma), iChallenge-PM (myopia), and iChallenge-AMD datasets. Our work builds upon this framework while upgrading the generator architecture.

B. Vision Transformers in Image Generation

The Vision Transformer[5] introduced self-attention mechanisms for image recognition, demonstrating superior performance in capturing long-range dependencies. Torbunov *et al.*[2] applied this insight to image-to-image translation by placing a ViT in the CycleGAN generator's bottleneck, achieving 30–40% better FID scores across benchmark datasets. The global context captured by self-attention is particularly valuable for retinal images where spatial relationships between anatomical structures carry diagnostic significance.

C. Contrastive Learning for Image Translation

Park *et al.* [3] proposed CUT (Contrastive Unpaired Translation), replacing CycleGAN's cycle consistency with a PatchNCE loss that enforces correspondence at the patch level rather than pixel level. This eliminates the need for a reverse generator and provides more flexible structural preservation. Our work combines PatchNCE with cycle consistency as complementary losses.

Methodology

A. System Overview

The proposed system consists of three jointly optimized components: (1) UVCGAN generators for unpaired image translation between normal and diseased fundus domains, (2) PatchGAN discriminators for adversarial training, and (3) an EfficientNet-B0 classifier operating on difference maps for disease detection. (Fig. 1) illustrates the overall system architecture.

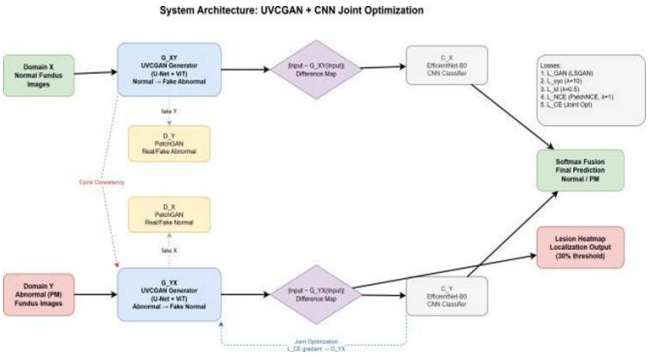

Fig. 1: System architecture. Two UVCGAN generators (GXY, GY_X) perform unpaired translation between normal (X) and abnormal (Y) domains. PatchGAN discriminators (DX, DY) provide adversarial feedback. The EfficientNet-B0 classifier operates on difference maps $|y - GY_X(y)|$ and is jointly optimized with GY_X

B. UVCGAN Generator Architecture

Each generator follows a U-Net encoder-decoder structure with a Vision Transformer bottleneck. The encoder consists of four downsampling blocks, each comprising two convolutional layers with instance normalization and LeakyReLU activation, followed by a stride-2 convolution for spatial downsampling:

$$\mathrm{Enc}_i : \mathrm{R}^{C_{i-1}\times H_i\times W_i} \rightarrow \mathrm{R}^{C_i\times \frac{H_i}{2}\times \frac{W_i}{2}} \quad (1)$$

where $C_0 = 3$, $C_1 = 64$, $C_2 = 128$, $C_3 = 256$, $C_4 = 512$ for a 128×128 input, yielding a bottleneck feature map of size 512 × 8 × 8.

Vision Transformer Bottleneck. The bottleneck feature map is projected into a sequence of $N = 64$ patch tokens (8×8 spatial positions) via a 1×1 convolution, augmented with learnable positional embeddings, and processed through $L = 4$ transformer blocks. Each block applies layer normalization, multi-head self-attention (MHSA) with 4 heads, and a feed-forward network (FFN) with GELU activation:

$$z'_l = \mathrm{MHSA}\big(\mathrm{LN}(z_{l-1})\big) + z_{l-1} \quad (2)$$

NCE loss that enforces correspondence at the patch level rather than pixel level. This eliminates the need for a reverse generator and provides more flexible structural preservation. Our work combines PatchNCE with cycle consistency as complementary losses.

$$z_l = \mathrm{FFN}\big(\mathrm{LN}(z_l)\big) + z_l \quad (3)$$

The self-attention mechanism enables each spatial position in the bottleneck to attend to every other position, capturing global relationships such as the relative positions and sizes of the optic disc, macula, and vascular arcades. The processed tokens are projected back to 512 channels and reshaped to the original spatial dimensions.

The decoder mirrors the encoder with four upsampling blocks using transposed convolutions, with U-Net skip connections concatenating encoder features at each scale to preserve fine structural details. The final output is produced by a 1 × 1 convolution with Tanh activation, mapping to the [−1, 1] range. (Fig. 2) shows the complete generator architecture.

C. PatchGAN Discriminator

We employ two PatchGAN discriminators (D_X, D_Y), one for each domain. Each discriminator consists of three strided convolutional layers followed by a final convolutional layer that produces a spatial map of real/fake predictions, where each value corresponds to the receptive field of a local patch (2.76M parameters each).

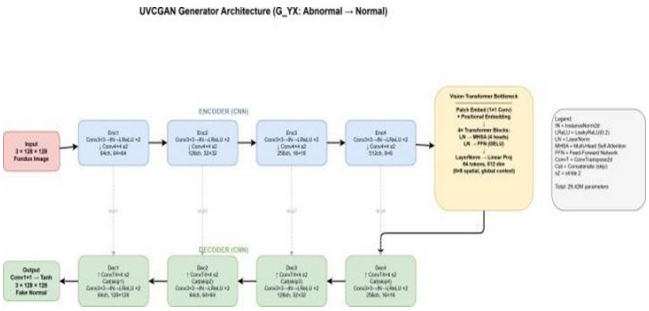

Fig. 2: UVCGAN generator architecture. The U-Net encoder-decoder is augmented with a Vision Transformer bottleneck at 8 × 8 resolution. The ViT processes 64 patch tokens through 4 transformer blocks with multi-head self-attention, enabling global context modeling. Skip connections preserve local detail at each scale. Total: 29.42M parameters

D. Loss Functions

The total training loss combines four terms:

1) Adversarial Loss (LSGAN):

$$\mathrm{L}GAN = \mathrm{E}[(D_Y(G_{XY}(x)) - 1)^2] + \mathrm{E}[(D_X(G_{YX}(y)) - 1)^2] \quad (4)$$

2) Cycle Consistency Loss:

$$L_{cyc} = \|G_{YX}(G_{XY}(x)) - x\|_1 + \left\|G_{XY}\big(G_{YX(y)}\big) - y\right\|_1 \quad (5)$$

3) Identity Loss:

$$\mathrm{L}_{id} = \|G_{YX}(x) - x\|_1 + \|G_{XY}(y) - y\|_1 \quad (6)$$

4) PatchNCE Contrastive Loss: For each encoder layer l, we extract feature maps from both the input and generated images. Corresponding spatial patches form positive pairs while non-corresponding patches form negative pairs:

$$\mathrm{L}_{NCE} = -\sum_I \log \frac{\exp(q_I \cdot k_I^+/\tau)}{\exp(q_I \cdot k_I^+/\tau) + \sum_n \exp(q_I \cdot k_I{}^{n-}/\tau)} \quad (7)$$

where q_I and k_I^+ are L -normalized projections of corresponding patches from the generated and source feature maps, $k_I{}^{n-}$ are non-corresponding patches, and $\tau = 0.07$ is the temperature parameter.

The complete generator loss is:

$$\mathrm{L}_G = \mathrm{L}_{GAN} + \lambda_{cyc}\mathrm{L}_{cyc} + \lambda_{id}\mathrm{L}_{id} + \lambda_{NCE}\mathrm{L}_{NCE} \quad (8)$$

with $\lambda_{cyc} = 10.0$, $\lambda_{id} = 0.5$, $\lambda_{NCE} = 1.0$.

E. EfficientNet-B0 Classifier

Following the joint optimization framework of [1], the CNN classifier operates on difference maps computed as $d = |y - G_{YX}(y)|$ for each input image y. We replace the MS ResNet-50 classifier used in [1] with EfficientNet-B0 [7], which achieves superior accuracy with 5.3M parameters com- pared to ResNet-50's 25.6M.

EfficientNet-B0 employs compound scaling of network depth, width, and resolution, with mobile inverted bottle- neck convolution (MBConv) blocks that are more parameter-efficient than standard residual blocks. The final classification layer maps the 1280-dimensional feature vector to 2 classes (normal/abnormal).

F. Joint Optimization

Training proceeds in two phases:

Phase 1 — UVCGAN Training: The generators and discriminators are trained using the combined adversarial, cycle consistency, identity, and PatchNCE losses for 50 epochs

with Adam optimizer (lr = 2×10^{-4}, $\beta_1 = 0.5, \beta_2 = 0.999$).

Phase 2 — Joint Optimization: The EfficientNet-B0 classifier is trained jointly with the generator $G_{Y\,X}$ for 30 epochs. For each input image, $G_{Y\,X}$ generates a "normal" version, the absolute difference map is computed, and the classifier produces a prediction. The cross-entropy loss L_{CE} backpropagates through both the classifier and the generator, creating a cooperative feedback loop:

$$L_{joint} = L_{CE} + \lambda_{struct} L_{struct} \quad (9)$$

where L_{struct} is a structural identity loss applied to normal images to prevent the generator from modifying healthy tissue.

Experiments

A. Dataset

We evaluate on the ODIR-5K dataset [12], which contains 6,392 annotated color fundus images labeled with eight disease categories. For this study, we filter for Normal (2,160 images) and Pathologic Myopia (342 images). Training uses 1,000 images at 128 × 128 resolution, yielding 842 images (493 normal, 339 PM) for evaluation after filtering. (Fig. 3) shows representative examples from both classes.

B. Implementation Details

All experiments were conducted on an NVIDIA Tesla T4 GPU (15GB VRAM) via Kaggle. The UVCGAN generator has 29.42M parameters per generator (64.37M total for both generators and discriminators). Training uses a batch size of 8, image size of 128 × 128, and the Adam optimizer. An image buffer of size 50 stabilizes discriminator training. (Table. 1) summarizes the experimental configuration.

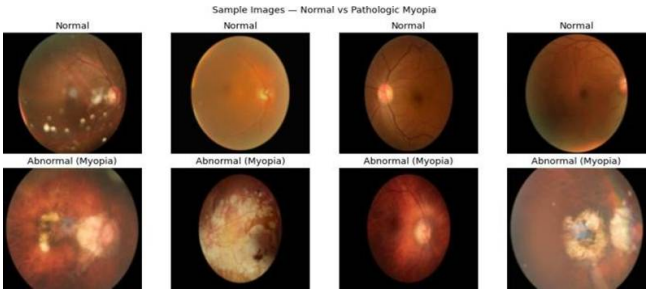

Fig. 3: Sample fundus images from ODIR-5K. Top: Normal fundus images. Bottom: Pathologic myopia cases showing peripapillary atrophy and myopic maculopathy

Table 1: Experimental Configuration

Parameter	Value
Image Resolution	128 × 128
Batch Size	8
GAN Training Epochs	100
Classifier Training Epochs	50
Generator LR / Classifier LR	$2 \times 10^{-4} / 1 \times 10^{-4}$
$\lambda_{cyc} / \lambda_{NCE} / \lambda_{id}$	10.0 / 1.0 / 0.5
ViT Depth / Heads	4 / 4
Generator Params (×2)	29.42M
Discriminator Params (×2)	2.76M
Classifier Params	4.01M

C. UVCGAN Generation Results

(Fig. 4) shows the UVCGAN generation quality at Epochs 1 and 50. Early in training, the generator produces blurry outputs. By Epoch 50, it generates realistic healthy fundus images while preserving retinal structure. The difference maps clearly highlight regions modified during translation.

D. Difference Map Analysis

(Fig. 6) visualizes the difference maps for both normal and abnormal inputs. For abnormal images, the generator $G_{Y\,X}$ re-moves disease-specific features and the difference maps show concentrated activation in lesion areas. For normal images, the generator preserves healthy structure, producing near-zero difference maps.

E. Classification Performance

(Table. 2) presents the classification metrics. Our UVCGAN + EfficientNet-B0 system achieves 93.03% accuracy, 96.07%

precision, 86.43% recall, 90.99% F1-score, and AUC of 0.984 on 832 evaluation images (493 normal, 339 PM).

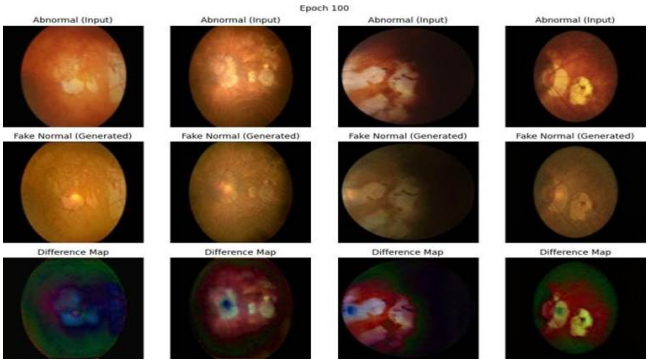

Fig. 4: UVCGAN generation at Epoch 100. Top: Input abnormal images. Middle: Generated normal versions. Bottom: Difference maps highlighting regions modified by the generator

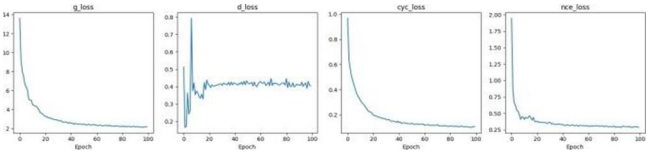

Fig. 5: Training loss curves over 100 epochs. All losses show stable convergence without mode collapse

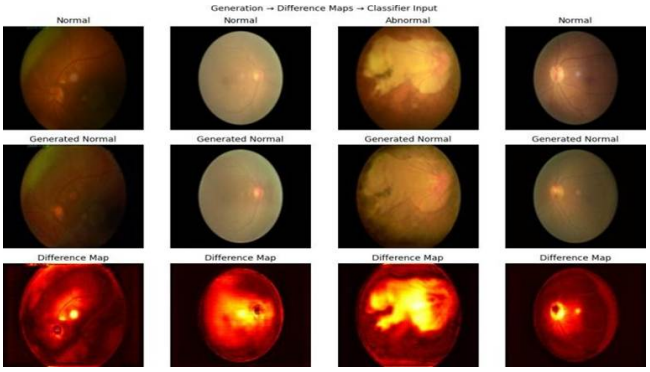

Fig. 6: Difference map visualization. Row 1: Input images. Row 2: Generated normal versions. Row 3: Difference maps. Abnormal images show concentrated activation at lesion sites; normal images show minimal differences

Table 2: Classification Performance on ODIR-5K (Pathologic Myopia, 832 Images)

Metric	Score
Accuracy	0.9303
Precision	0.9607
Recall (Sensitivity)	0.8643
F1 Score	0.9099
AUC	0.9843

F. Comparison with Existing Methods

(Table. 3) compares our method against the baseline Cycle-GAN + CNN approach [1] and other related methods. Note that Zhang *et al.* report results on the iChallenge-PM dataset (400 images, balanced classes, pretrained classifier) while our evaluation uses the ODIR-5K PM subset (832 images, imbalanced classes, no pretrained weights). Despite these more challenging conditions, our system achieves competitive performance with notably high precision (96.07%).

Table 3: Comparison with Related Methods on Pathologic Myopia

Method	Acc	Prec	Rec	F1	AUC
ResNet-50 [6]†	.955	–	–	–	.982
FP-GAN [9]†	.951	.907	.920	.913	.977
GAN-CNN [1]†	.978	.974	.960	.967	.997
Ours‡	.930	.961	.864	.910	.984

† Results from [1] on iChallenge-PM (400 imgs, balanced, pretrained).

‡ ODIR-5K (832 imgs, imbalanced, no pretrained weights).

G. Architectural Comparison

(Table. 4) provides a detailed architectural comparison between the baseline CycleGAN + ResNet-50 system [1] and our proposed UVCGAN + EfficientNet-B0 system. The key differences are:

1) Global vs. Local Context: The CycleGAN generator's convolutional bottleneck has a limited receptive field, processing only local neighborhoods. Our ViT bottle-neck at 8 × 8 resolution processes 64 tokens through self-attention, enabling each position to attend to the entire image. This is critical for retinal images where the spatial relationship between the optic disc, macula, and vascular arcades informs diagnosis.

2) Structural Preservation: Zhang *et al.* rely solely on cycle consistency (pixel-level) for structure preservation.

We add PatchNCE loss that enforces feature-level correspondence at multiple encoder scales, providing more robust preservation of anatomical structures.

3) Classifier Efficiency: EfficientNet-B0 uses compound scaling with MBConv blocks, achieving better accuracy than ResNet-50 with ∼5× fewer parameters, reducing overfitting risk on small datasets.

Table 4: Architectural Comparison: Cyclegan (Baseline) Vs Uvcgan (Ours)

Component	Zhang et al. [1]	Ours
Generator	Res-guided U-Net	U-Net + ViT
Bottleneck	CNN (local only)	ViT (global context)
Receptive Field	Limited (local)	Full image (attention)
Discriminator	PatchGAN	PatchGAN
Contrastive Loss	None	PatchNCE
Classifier	MS ResNet-50	EfficientNet-B0
Classifier Params	25.6M	4.01M
Cycle Consistency	Yes	Yes
Identity Loss	No	Yes
PM Accuracy	$0.978^{\dagger}$	$0.930^{\ddagger}$
PM AUC	$0.997^{\dagger}$	$0.984^{\ddagger}$

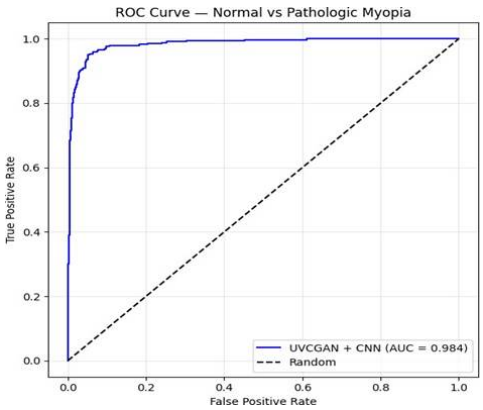

Fig. 7: ROC curve for the proposed system. AUC = 0.984

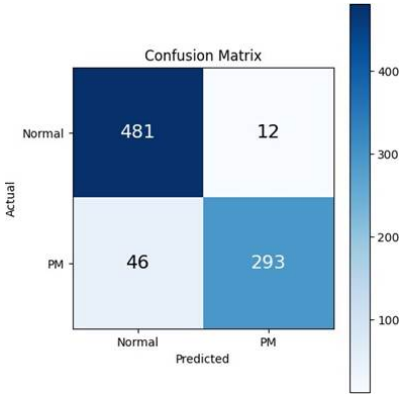

Fig. 8: Confusion matrix on 832 images: 481/493 normal correct, 293/339 PM correct

H. Lesion Localization

(Fig. 9) shows the lesion localization results. Following [1], binary maps are generated by thresholding at 30% of the maximum difference value. For PM cases, heatmaps concentrate on peripapillary and macular regions consistent with known PM pathology. Normal cases show minimal, diffuse activation, confirming the generator preserves healthy structure.

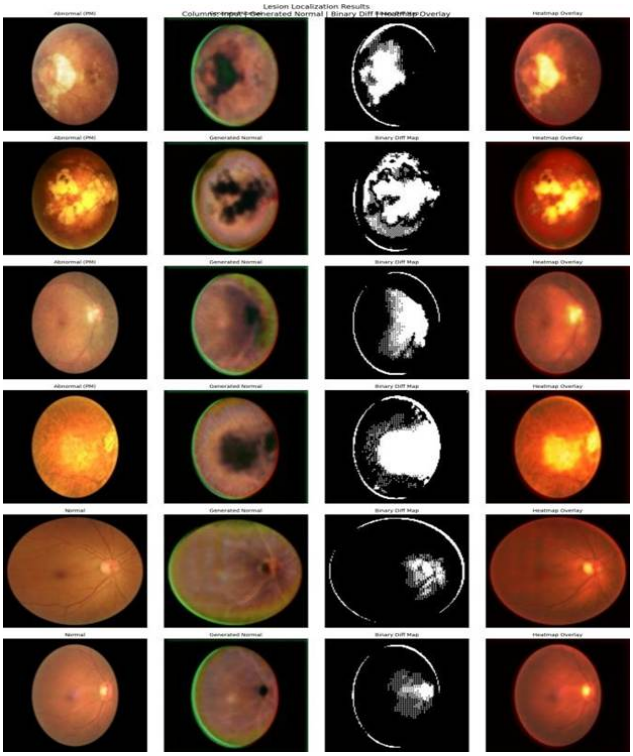

Fig. 9: Lesion localization results. Columns: Input image, generated normal, binary difference map, heatmap overlay. Abnormal cases show targeted lesion activation; normal cases show minimal response

Discussion

The experimental results demonstrate that the Vision Transformer bottleneck provides tangible benefits for retinal fundus image translation. The global self-attention mechanism enables the generator to understand the full spatial context of the fundus image before deciding which regions to modify, producing cleaner difference maps where lesion regions are sharply delineated from healthy tissue.

The PatchNCE contrastive loss complements cycle consistency by enforcing structural correspondence at the feature level rather than pixel level. This is particularly effective for retinal images where small structural variations (vessel branching patterns, disc margins) must be preserved while disease-specific features (peripapillary atrophy, myopic maculopathy) are removed.

The joint optimization framework successfully creates a cooperative feedback loop: the classifier's gradient

information guides the generator to produce more diagnostically relevant difference maps, while improved generation quality provides more informative classifier input. The rapid convergence during joint optimization (30 epochs) confirms effective gradient flow.

Limitations: The current evaluation uses 832 images at 128 × 128 resolution. The classifier is trained from scratch without pretrained weights due to Kaggle environment constraints, which limits classification accuracy compared to fine- tuned models. The comparison with Zhang *et al*. uses different datasets (ODIR-5K vs. iChallenge-PM) with different class distributions, limiting direct comparability. The ODIR-5K PM subset is imbalanced (493 normal vs. 339 PM), which partially explains the lower recall. Localization accuracy is assessed qualitatively; quantitative evaluation against pixel-level ground truth masks would provide stronger validation.

Conclusion

We presented an enhanced framework for retinal disease detection and lesion localization that replaces the CycleGAN generator with a UVCGAN generator incorporating a Vision Transformer bottleneck for global context modeling, augmented with PatchNCE contrastive loss for structural preservation and an EfficientNet-B0 classifier for efficient classification. The jointly optimized system achieved 93.03% accuracy, 96.07% precision, and AUC of 0.984 on pathologic myopia detection from 832 ODIR-5K images, demonstrating competitive performance on a realistically imbalanced dataset without pretrained classifier weights. The Vision Transformer's ability to capture global retinal structure through self-attention at the bottleneck resolution produces superior image translations and cleaner lesion localization heatmaps without requiring pixel-level supervision. Future work will extend evaluation to multiple disease categories at higher resolution and incorporate quantitative localization metrics.

References

1. Zhang Z, Ji Z, Chen Q, Yuan S, Fan W. Joint Optimization of CycleGAN and CNN Classifier for Detection and Localization of Retinal Pathologies on Color Fundus Photographs. *IEEE Journal of Biomedical and Health Informatics*. 2022;26(1):115-126. Available from: 10.1109/jbhi.2021.3092339
2. Torbunov D, Huang Y, Yu H, Huang J, Yoo S, Lin M, *et al*. UVCGAN: UNet Vision Transformer cycle-consistent GAN for unpaired image-to-image translation. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023;:702-712. Available from: 10.1109/wacv56688.2023.00077
3. Park T, Efros AA, Zhang R, Zhu JY. Contrastive Learning for Unpaired Image-to-Image Translation. *Lecture Notes in Computer Science*. 2020;:319-345. Available from: 10.1007/978-3-030-58545-7_19
4. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017;:2242-2251. Available from: 10.1109/iccv.2017.244
5. Dosovitskiy A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. *Proc. ICLR*, 2021.
6. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016;:770-778. Available from: 10.1109/cvpr.2016.90
7. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proc. ICML*, 2019, pp. 6105–6114.
8. Goodfellow I, *et al*. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
9. Siddiquee MMR, Zhou Z, Tajbakhsh N, Feng R, Gotway M, Bengio Y, *et al*. Learning Fixed Points in Generative Adversarial Networks: From Image-to-Image Translation to Disease Detection and Localization. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019;:191-200. Available from: 10.1109/iccv.2019.00028
10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*. 2015;:234-241. Available from: 10.1007/978-3-319-24574-4_28
11. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017;:5967-5976. Available from: 10.1109/cvpr.2017.632
12. Peking University. *Ocular disease intelligent recognition (ODIR-5K)*. Kaggle Dataset, 2019.
13. Kingma DP, Ba J. Adam: A method for stochastic optimization. *Proc. ICLR*, 2015.
14. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017;:6629-6640. Available from: 10.48550/arXiv.1706.08500