Progressive Optimization of HydraLA-Net for Microaneurysm Segmentation

Jessica Yuan
University of Waterloo
jessica.yuan1@uwaterloo.ca

Michael Liu
University of Waterloo
m529liu@uwaterloo.ca

Andrew Yang
University of Waterloo
a289yang@uwaterloo.ca

William Chiu
University of Waterloo
w4chiu@uwaterloo.ca

Sidharth Shah
University of Waterloo
s78shah@uwaterloo.ca

Tom Almog
University of Waterloo
talmog@uwaterloo.ca

Apisan Kaneshan
University of Waterloo
askanesh@uwaterloo.ca

Christopher Risi
University of Waterloo
christopher.risi@uwaterloo.ca

***Abstract*—Microaneurysms are the earliest detectable sign of diabetic retinopathy, yet automated segmentation remains challenging due to their small size, low contrast, and severe class imbalance in fundus images. In this work, we extend the Lesion-Aware Network (LA-Net) with class-specific prediction heads to reduce gradient competition during training. We conduct an experimental study on preprocessing techniques, including CLAHE variants, and imbalance-aware loss functions using a progressive optimization strategy across three public datasets: IDRiD, DDR, and TJDR. Results demonstrate improved microaneurysm segmentation while maintaining competitive performance on other lesion classes, providing a practical framework for enhancing early diabetic retinopathy detection. A full implementation is available at https://github.com/jessicayuan1/fundus-image-segmentation.**

I. Introduction

Medical image segmentation has become a cornerstone of computer-aided diagnosis, enabling clinicians to detect and monitor disease progression with greater precision and efficiency. In ophthalmology, automated analysis of retinal fundus images offers particular promise: with over 537 million adults living with diabetes worldwide and approximately one-fifth [1] developing diabetic retinopathy, scalable screening solutions are urgently needed to prevent irreversible vision loss. Yet despite significant advances in deep learning architectures, reliable detection of the earliest and most subtle retinal lesions remains an open challenge.

A. Motivation

Diabetic retinopathy is the leading cause of preventable blindness. Among diabetic retinopathy lesions, microaneurysms hold particular clinical significance as they are the earliest detectable sign of the disease [2], and their presence of even a few microaneurysms is sufficient to diagnose mild nonproliferative diabetic retinopathy [3].

However, accurate segmentation of microaneurysms remains exceptionally difficult. These lesions are extremely small, often only a few pixels in diameter, and exhibit low contrast against surrounding retinal tissue. Their subtle appearance makes them difficult to distinguish from blood vessels or imaging noise, limiting the reliability of automated systems for early detection [4].

B. Related Works

Segmentation Architectures

U-Net, introduced by Ronneberger et al. [5], was the standard architecture for medical image segmentation. The standard U-Net treats all parts of the image equally, which limits its ability to focus on small structures, which are common in our task. Oktay et al. addressed this with Attention U-Net [6], which learns to focus on relevant regions while suppressing irrelevant ones. For diabetic retinopathy specifically, Xia et al. proposed LA-Net [7], a lesion-aware architecture designed to better preserve and emphasize lesion-related features during learning. We adapt this architecture for our HydraLA-Net.

Color Channels and Contrast Enhancement

Biswas et al. investigated which color channel is best for diagnosing retinal diseases, finding that the green channel provides the best contrast among individual channels and the least noise for tasks like vessel segmentation. The red and blue channels have been shown to be prone to contrast, noise, and inappropriate exposure. Nonetheless, Biswas et al. emphasize that all color channels play an important role in fundus photographs [8].

Loss Functions for Imbalanced Segmentation

Dice loss has been widely used for imbalanced segmentation by optimizing region overlap rather than pixel-wise accuracy [9]. However, it equally penalizes false positives and false negatives, which is not optimal for lesion detection tasks where missed lesions are more critical. To address this limitation, Abraham and Khan introduced Focal Tversky loss, which allows asymmetric penalization between false positives and false negatives [10].

C. Problem Definition

Automated DR lesion segmentation requires identification of four lesion classes: microaneurysms, hemorrhages, soft exudates, and hard exudates. Among these, microaneurysms are the most challenging due to their small size, low contrast, and severe class imbalance.

Microaneurysms occupy only a small fraction of each image, while background pixels dominate, biasing standard training toward the majority class. Their few-pixel diameter

and subtle appearance further complicate discrimination from vessels or imaging noise [11].

This difficulty is reflected in prior work. For example, on IDRiD, LA-Net achieves an F_1 score of 0.243 for microaneurysms compared to 0.626 for hard exudates [7].

D. Our Contributions

In our work, we extend LA-Net into HydraLA-Net and conduct a series of experimental studies to improve **multi-label** diabetic retinopathy lesion segmentation. Our approach is a **progressive optimization** strategy, where each variable in each part is evaluated and the best-performing configuration, if one is obvious, is carried forward to the next stage.

Specifically, we evaluate architectural modifications, preprocessing strategies (CLAHE variants), loss functions for class imbalance, and provide a dataset-wise analysis of performance.

All experiments are conducted in a **multi-label** setting, where the model is trained to segment all four lesion types simultaneously as four separate binary problems. Joint multi-label training enables the model to leverage visual characteristics shared across lesion classes while preserving class-specific predictions.

As such, our experimental setup greatly differs from the original LA-Net paper, and thus direct numerical comparisons with the original LA-Net results are not appropriate.

II. Methodology

A. Datasets

We conduct our study on three publicly available fundus datasets: IDRiD, DDR, and TJDR. The first two provide binary pixel-level annotations for four lesion types, while TJDR provides a color-coded annotation map that we decompose into four binary masks.

IDRiD contains 81 segmentation images with corresponding lesion masks [12]. **DDR** includes 13,673 images, of which 757 provide segmentation masks [13]. **TJDR** consists of 561 high-resolution fundus images with color-coded lesion annotations [14].

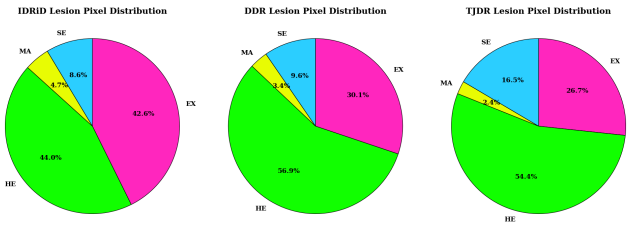

Fig. 1. Dataset Pixel Distributions by Lesion Class
Color legend: Green = Hemorrhages (HE); Yellow = Microaneurysms (MA); Blue = Soft Exudates (SE); Pink = Hard Exudates (EX).

The above figure shows the class imbalance in terms of the number of pixels per lesion class. This is common for DR datasets. This imbalance can bias learning against underrepresented classes.

The IDRiD and TJDR datasets provide predefined training and testing splits. We reserve the official testing sets for final evaluation and perform an 80/20 split on the training sets to obtain training and validation subsets. The DDR dataset includes predefined training, validation, and testing splits. We exclude the testing set and combine the training and validation sets, followed by an 80/20 split to construct the final training and validation subsets.

For the individual dataset analysis, we make the same 80/20 split on the training sets, but for each individual dataset. The testing sets remain the one provided by each dataset.

TABLE I
Dataset Split Sizes

Dataset	Training Set	Validation Set	Testing Set
IDRiD	43	11	27
DDR	426	106	225
TJDR	358	90	113
Total	**827**	**207**	**365**

B. Model Architecture

Our model is based on the Lesion-Aware (LA-Net) Network [7] which featured a ResNet-50 encoder with pretrained Image-Net weights, and a Lesion-Aware Module (LAM) to aid in the capturing of small details. The LAM works by utilizing orientation-aware features and a self-attention mechanism to direct focus to small lesions. LA-Net also contains a Feature Preserve Module (FPM), containing a Feature Preserve Block (FPB) and Feature Fusion Block (FFB) to maintain lesion information and the fusion of global-disease related information, respectively. The FPB maintains multi-scale features and preserves information from the encoder. The FFB progressively aggregates features from the network and from the LAM [7].

In our paper, we propose the addition of a Hydra Head that splits the final segmentation head into four class-specific prediction branches, which prevents some competition between gradients, increasing sensitivity towards microaneurysms. Each head consists of a 1×1 convolution, GELU activation, and a final 1×1 convolution producing per-pixel class probabilities.

C. Data Preprocessing and Augmentation

CLAHE (Contrast Limited Adaptive Histogram Equalization) and its variants are applied as deterministic preprocessing rather than augmentation. When used, transformations are applied to all training, validation, and test images for consistent evaluation.

CLAHE enhances local contrast through tile-wise histogram equalization with clipping to limit noise amplification. We ablate the clip limit parameter to assess contrast strength [15].

We evaluate three CLAHE-based preprocessing variants:

- CLAHE-L: CLAHE applied to the L channel in LAB space, followed by conversion back to RGB.
- CLAHE-G: CLAHE applied to the green channel only.

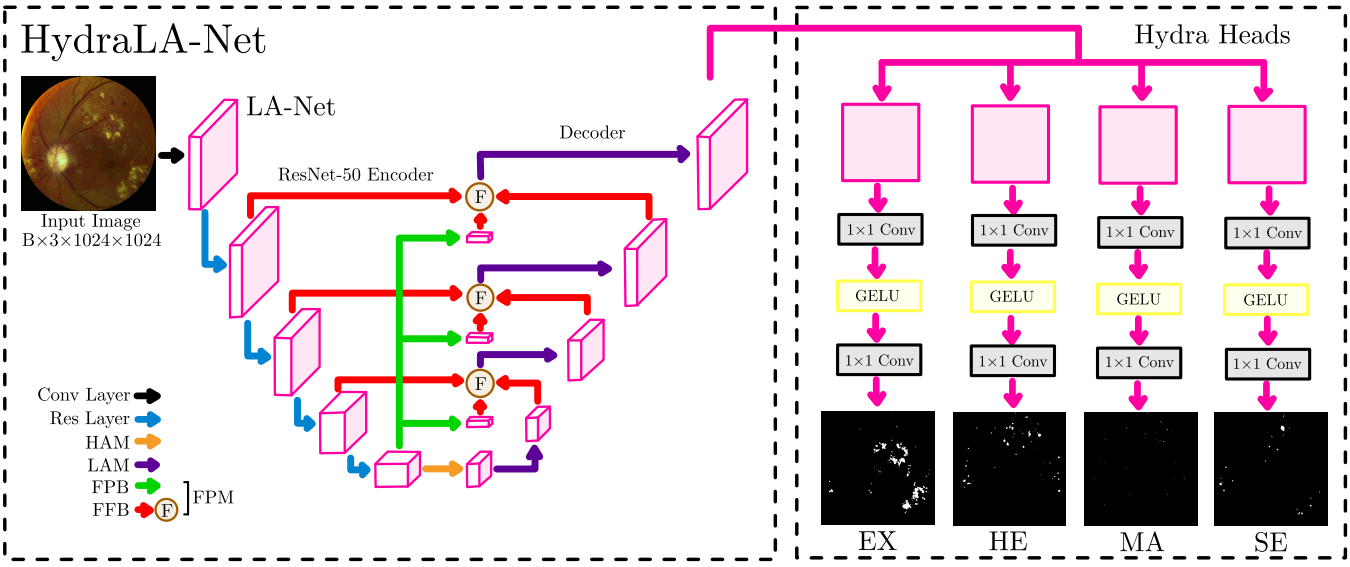

Fig. 2. HydraLA-Net Model Flowchart

- CASP: CLAHE on the green channel with Gaussian blur applied to red and blue channels in an attempt to reduce noise.

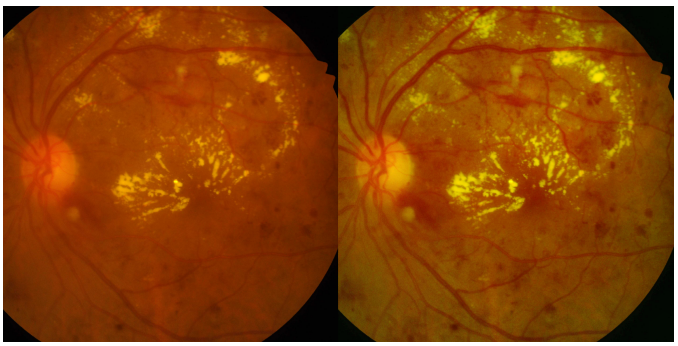

Fig. 3. Left: No CLAHE, Right: CLAHE-G

Furthermore, standard probabilistic augmentations (flips, rotations, zoom, spatial shifts) are applied during training only. Lastly, all pixel values are normalized using ImageNet means and standard deviations.

D. Loss Functions

Binary Cross-Entropy (BCE) is a widely used loss function for binary tasks and is commonly applied to semantic segmentation problems where each pixel is treated as an independent binary prediction.

$$\mathcal{L}_{\text{BCE}} = -\left[y \log(p) + (1-y)\log(1-p)\right] \tag{1}$$

$y \in \{0, 1\}$ denotes the ground-truth
$p \in [0, 1]$ denotes the predicted probability

BCE measures how well predicted probabilities align with ground-truth labels and strongly penalizes confident but incorrect predictions. As a result, BCE encourages well-calibrated predictions by discouraging overconfident errors during training. However, BCE does not explicitly address severe foreground-background imbalance, which can bias optimization toward the dominant background class in lesion segmentation [16].

Dice Loss optimizes region overlap and is commonly used for imbalanced segmentation. However, it assigns equal weight to false positives and false negatives, which is suboptimal when missed lesions are more critical [10].

The Tversky index (TI) improves on Dice loss by introducing two hyperparameters, α and β, to control the penalties on FP's and FN's, respectively. Setting $\beta > \alpha$ makes it effective for tasks where false negatives are a greater concern than false positives. The TI is converted into Tversky loss (TL) by minimizing 1 - TI. When $\alpha = \beta$, TL simplifies to dice loss. By default, when α and β are scalar values, TL is not class-imbalance aware. Explicit class-specific imbalance can only be addressed by assigning different α/β parameters per class as vectors [10].

$$\text{TI}^{(c)} = \frac{|\text{TP}^{(c)}|}{|\text{TP}^{(c)}| + \alpha|\text{FP}^{(c)}| + \beta|\text{FN}^{(c)}|} \tag{2}$$

$|TP|$ denotes the number of true positive pixels
$|FP|$ denotes the number of false positive pixels
$|FN|$ denotes the number of false negative pixels
c denotes a specific lesion class

Focal Tversky Loss (FT) further extends on Tversky loss by introducing a focusing parameter γ, to emphasize hard,

misclassified pixels. In this study, we use only a scalar value for γ and use a mix of scalar and vector values for α/β.

$$L_{FT} = \frac{1}{C}\sum_{c=1}^{C}\left(1 - \mathrm{TI}^{(c)}\right)^{\gamma} \tag{3}$$

C denotes the number of classes
γ denotes the focusing parameter
$\mathrm{TI}^{(c)}$ denotes the Tversky Index for lesion class c

For our study, we define the following loss functions (γ is held constant for all loss functions at 2.0):

$$\mathcal{L}_1 = 0.8\mathcal{L}_{FT} + 0.2\mathcal{L}_{BCE} \quad \alpha = 0.5 \quad \beta = 0.5 \tag{4}$$

$$\mathcal{L}_2 = 0.8\mathcal{L}_{FT} + 0.2\mathcal{L}_{BCE} \quad \alpha = 0.4 \quad \beta = 0.6 \tag{5}$$

$$\begin{aligned} \mathcal{L}_3 &= 0.8\mathcal{L}_{FT} + 0.2\mathcal{L}_{BCE} \\ \alpha &= [0.4,\ 0.4,\ 0.2,\ 0.4], \\ \beta &= [0.4,\ 0.4,\ 0.2,\ 0.4] \end{aligned} \tag{6}$$

$$\begin{aligned} \mathcal{L}_4 &= 1.0\mathcal{L}_{FT} \\ \alpha &= [0.4,\ 0.4,\ 0.2,\ 0.4], \\ \beta &= [0.4,\ 0.4,\ 0.2,\ 0.4] \end{aligned} \tag{7}$$

E. Control Variables

All legs of the experiment feature the HydraLA-Net model architecture (baseline also includes LA-Net), and constant data (with the exception of the individual dataset analysis) from the IDRiD, DDR, and TJDR datasets. Augmentations and preprocessing (based on the given experiment) remain fixed. In training, we use the Adam Optimizer with learning rate (1e-5) and weight decay (1e-4). Dataloaders use a constant batch size of 2. All models are trained for a fixed number of 150 epochs. The ResNet-50 backbone is initialized with ImageNet-pretrained weights. The input resolution of images is held constant at 1024x1024 pixels. Finally, the CLAHE grid size is kept constant at 8x8 pixels.

F. Evaluation Metrics

We evaluate segmentation performance using recall and F_1 scores.

Recall is defined as the ratio of true positives to the total number of actual positive instances.

$$\mathrm{Recall} = \frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FN}|} \tag{8}$$

Recall alone is insufficient to evaluate segmentation performance, as a model that predicts all pixels as positive would achieve perfect recall but extremely low precision. The F_1 score, defined as the harmonic mean of precision and recall, balances sensitivity with predictive accuracy. As a result, the F_1 score provides a more comprehensive evaluation metric for imbalanced segmentation tasks, and we adopt it as our primary evaluation metric.

$$\mathrm{F}_1 = \frac{2|\mathrm{TP}|}{2|\mathrm{TP}| + |\mathrm{FP}| + |\mathrm{FN}|} \tag{9}$$

G. Evaluation Protocol

For each experiment, models are trained on the training split and monitored using the validation split. The best checkpoint is selected based on the highest averaged validation F_1 score and is subsequently evaluated once on the held-out test set. The testing set is not used during training or model selection.

Model performance is evaluated using the recall and F_1 scores. We report the macro-average F_1 score, as well as per class. For recall, we report only the macro-average recall, as well as recall for the microaneurysm class. We display all scores multiplied by 100.

H. Experimental Design

We conduct four experiments following a progressive optimization strategy.

- **Architecture Comparison**: LA-Net and HydraLA-Net are trained under the same multi-label configuration without CLAHE preprocessing, using L_1, totalling two configurations.
- **Preprocessing Ablation**: HydraLA-Net is evaluated with three CLAHE variants (CLAHE-L, CLAHE-G, CASP) and two clip limits (1.5, 2.5), totaling six configurations.
- **Loss Function Ablation**: The two best-performing CLAHE variants and the best-performing clip limit from the previous stage, if a clear performance advantage is observed, are retained. We experiment using L_2, L_3, and L_4, totalling six configurations.
- **Individual Dataset Analysis**: We evaluate each dataset (IDRiD, DDR, TJDR) with and without CLAHE to assess dataset-specific effects, using the best performing CLAHE variant and loss function, totalling six configurations.

III. Results & Discussion

TABLE II
BASELINE PERFORMANCE

Model	F_1					Recall	
	Mean	EX	HE	MA	SE	Mean	MA
HydraLA-Net	**47.5**	**61.8**	**51.1**	**31.4**	45.6	47.7	**39.2**
LA-Net	46.9	61.0	49.0	31.2	**46.4**	**48.6**	37.7

The results indicate that the addition of the Hydra Head module to LA-Net provides a small boost in MA recall from $(37.7 \rightarrow 39.2)$, while overall F_1 performance remains largely similar to the base model. Performance on other lesion types is either slightly improved or maintained, with the largest gain observed for hemorrhages. This aligns with the theoretical basis of the segmentation head module, with its purpose being to diminish competition between lesions.

TABLE III
CLAHE PREPROCESSING VARIATIONS

CLAHE Variant	F_1					Recall	
	Mean	EX	HE	MA	SE	Mean	MA
CLAHE-L							
CL = 1.5	48.2	61.5	51.3	34.1	45.9	48.3	37.2
CL = 2.5	47.7	62.7	48.7	32.5	46.7	46.6	35.2
CLAHE-G							
CL = 1.5	**49.3**	**63.6**	50.2	**34.3**	49.0	**50.1**	40.6
CL = 2.5	47.3	62.4	49.1	33.3	44.5	47.6	37.2
CASP							
CL = 1.5	**49.3**	61.0	**53.2**	33.8	49.0	46.8	40.5
CL = 2.5	47.2	60.7	48.5	29.4	**50.3**	48.4	**43.3**

Targeted contrast enhancement through CLAHE demonstrates measurable improvements in segmentation performance, with green channel preprocessing (CLAHE-G) achieving the highest mean F_1 of 49.3 and best microaneurysm F_1 of 34.3 at clip limit 1.5. This aligns with prior findings by Biswas et al., who reported that the green channel provides superior lesion contrast while red and blue channels are prone to oversaturation and noise.

CASP preprocessing achieves an equivalent mean F_1 (49.3) and the highest hemorrhage score (53.2), suggesting that selective channel preprocessing offers complementary benefits across lesion types.

Increasing the clip limit to 2.5 results in a noticeable overall decline in performance across all variants. For example, CLAHE-G shows a decrease in microaneurysm F_1 from $(34.3 \rightarrow 33.3)$, while CASP exhibits a larger drop from $(33.8 \rightarrow 29.4)$. This suggests that overly aggressive contrast enhancement may amplify noise and reduce segmentation stability. Based on these results, CLAHE-G and CASP with a clip limit of 1.5 are selected for subsequent experiments.

TABLE IV
LOSS FUNCTION COMPARISON

Loss Function	F_1					Recall	
	Mean	EX	HE	MA	SE	Mean	MA
CLAHE-G							
L_2	49.4	63.2	51.8	**33.4**	49.1	51.4	47.9
L_3	**50.1**	**65.1**	50.4	**33.4**	**51.5**	51.6	47.8
L_4	48.4	62.6	50.3	30.8	49.8	47.5	42.5
CASP							
L_2	49.0	62.0	51.5	33.2	49.2	48.4	42.7
L_3	47.8	62.9	**52.9**	29.8	45.7	**55.4**	**53.1**
L_4	48.0	63.6	50.1	30.4	48.0	51.1	47.9

The results indicate that class-imbalance-aware loss functions improve HydraLA-Net performance, although effectiveness depends on the preprocessing strategy. Under CLAHE-G, L_3 achieves the highest mean F_1 of 50.1, improving upon L_2 $(49.4 \rightarrow 50.1)$. Both L_2 and L_3 achieve equivalent microaneurysm F_1 scores of 33.4, while L_4, which removes BCE entirely, shows a noticeable decline in microaneurysm F_1 $(33.4 \rightarrow 30.8)$, confirming that BCE serves an important regularization role.

Under CASP preprocessing, L_3 achieves the highest microaneurysm recall $(42.7 \rightarrow 53.1)$ but at the cost of reduced microaneurysm F_1 $(33.2 \rightarrow 29.8)$, suggesting that higher beta values successfully increase sensitivity while introducing additional false positives. This aligns with the theoretical basis of Focal Tversky Loss as described by Abraham and Khan. However, L_4 does not recover performance, further supporting the importance of retaining BCE in the loss formulation.

Overall, L_3 with green channel CLAHE provides the most balanced performance, validating that class-imbalance aware loss functions combined with BCE offer meaningful improvements for multi-label diabetic retinopathy lesion segmentation.

TABLE V
INDIVIDUAL DATASET ANALYSIS

Dataset & CLAHE	F_1					Recall	
	Mean	EX	HE	MA	SE	Mean	MA
IDRiD							
CLAHE Off	57.2	70.8	54.2	39.5	64.2	62.5	57.7
CLAHE-G	58.6	74.0	54.5	39.7	66.4	61.9	64.2
DDR							
CLAHE Off	38.8	56.2	41.5	22.6	34.9	44.8	40.0
CLAHE-G	42.7	59.4	45.4	24.7	41.4	46.9	46.9
TJDR							
CLAHE Off	56.2	61.9	67.8	41.4	53.9	62.1	51.5
CLAHE-G	56.6	63.3	67.8	37.7	57.8	61.8	42.7

$L3$ is selected as the loss function for all 6 configurations. Analysis across individual datasets reveals substantial variation in model performance, with IDRiD achieving the highest and most consistent results. With CLAHE green channel preprocessing, IDRiD attains a mean F_1 of 58.6, microaneurysm F_1 of 39.7, and the highest microaneurysm recall of 64.2. This consistency likely reflects the dataset's controlled acquisition conditions and high-quality expert annotations from a single clinical site.

DDR exhibits the weakest performance across all metrics, achieving only 42.7 mean F_1 and 24.7 microaneurysm F_1, even with CLAHE preprocessing. This degradation likely stems from the dataset's heterogeneous nature, comprising of images from 147 hospitals with varying equipment and protocols. Nevertheless, CLAHE provides the largest relative improvement on DDR, increasing microaneurysm recall by 6.9 points, suggesting that contrast enhancement partially mitigates image quality inconsistencies.

TJDR presents an interesting case where CLAHE improves most metrics but reduces microaneurysm F_1 from $(41.4 \rightarrow 37.7)$, potentially due to label noise introduced during decomposition of its color-coded annotation scheme. These findings underscore the importance of dataset quality and standardization for robust microaneurysm segmentation.

IV. CONCLUSION

Overall, the results of this study demonstrate that targeted architectural modifications, preprocessing strategies, and loss function design collectively improve microaneurysm segmentation in diabetic retinopathy detection. Across the full-dataset

experiments, progressive optimization from Part 1 to Part 3 increased mean F_1 by 3.2 points $(46.9 \rightarrow 50.1)$, microaneurysm F_1 by 3.2 points $(31.2 \rightarrow 34.4)$, and microaneurysm recall by 15.4 points $(37.7 \rightarrow 53.1)$. The proposed HydraLA-Net, which introduces class-specific prediction heads to reduce gradient competition, achieves improved sensitivity to microaneurysms while maintaining competitive performance on larger lesion classes.

Our preprocessing ablation confirms that CLAHE-G at a moderate clip limit of 1.5 provides optimal contrast enhancement without introducing excessive noise, aligning with prior findings on fundus image analysis. Furthermore, class-imbalance-aware Focal Tversky loss combined with BCE offers the most balanced performance, validating the importance of asymmetric penalization for imbalanced segmentation tasks.

The individual dataset analysis reveals that performance varies substantially across datasets, with standardized acquisition protocols and high-quality annotations yielding the most consistent results. These findings highlight that robust microaneurysm segmentation depends not only on model architecture but also on dataset quality.

While microaneurysm segmentation remains challenging due to inherent class imbalance and lesion characteristics, our progressive optimization framework provides a practical methodology for systematically improving detection performance. Future work may explore additional architectural enhancements, such as multi-scale attention mechanisms, and investigate domain adaptation techniques to improve generalization across heterogeneous clinical datasets. Ultimately, improving microaneurysm detection supports earlier diagnosis of diabetic retinopathy and contributes to reducing preventable vision loss among diabetic patients.

V. Acknowledgements

We thank Dr. Nadine Furtado at the University of Waterloo School of Optometry and Vision Sciences for guidance in mapping the TJDR dataset's color-coded annotations to lesion classes. We also thank Rebecca Hutchinson, Liaison Librarian for Computer Science at the University of Waterloo Library, for her support with literature research and citation practices.

References

[1] Z. L. Teo, Y.-C. Tham, M. Yu, M. L. Chee, T. H. Rim, N. Cheung, M. M. Bikbov, Y. X. Wang, Y. Tang, Y. Lu, I. Y. Wong, D. S. W. Ting, G. S. W. Tan, J. B. Jonas, C. Sabanayagam, T. Y. Wong, and C.-Y. Cheng, "Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis," *Ophthalmology*, vol. 128, no. 11, pp. 1580–1591, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0161642021003213

[2] S. S. Feman, "The natural history of the first clinically visible features of diabetic retinopathy," *Transactions of the American Ophthalmological Society*, vol. 92, pp. 745–773, 1994.

[3] L. Wu, "Classification of diabetic retinopathy and diabetic macular edema," *World Journal of Diabetes*, vol. 4, no. 6, p. 290, 2013. [Online]. Available: http://www.wjgnet.com/1948-9358/full/v4/i6/290.htm

[4] L. Zhang, S. Feng, G. Duan, Y. Li, and G. Liu, "Detection of Microaneurysms in Fundus Images Based on an Attention Mechanism," *Genes*, vol. 10, no. 10, p. 817, Oct. 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6827155/

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, arXiv:1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

[6] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," May 2018, arXiv:1804.03999. [Online]. Available: http://arxiv.org/abs/1804.03999

[7] X. Xia, K. Zhan, Y. Fang, W. Jiang, and F. Shen, "Lesion-aware network for diabetic retinopathy diagnosis," Aug. 2024, arXiv:2408.07264. [Online]. Available: http://arxiv.org/abs/2408.07264

[8] S. Biswas, M. I. A. Khan, M. T. Hossain, A. Biswas, T. Nakai, and J. Rohdin, "Which Color Channel Is Better for Diagnosing Retinal Diseases Automatically in Color Fundus Photographs?" *Life*, vol. 12, no. 7, p. 973, Jun. 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9321111/

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings of the Fourth International Conference on 3D Vision (3DV)*, 2016.

[10] N. Abraham and N. M. Khan, "A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation," Oct. 2018, arXiv:1810.07842. [Online]. Available: http://arxiv.org/abs/1810.07842

[11] J. M. Stewart, M. Coassin, and D. M. Schwartz, "Diabetic Retinopathy," in *Endotext*, K. R. Feingold, R. A. Adler, S. F. Ahmed, B. Anawalt, M. R. Blackman, G. Chrousos, E. Corpas, W. W. de Herder, K. Dhatariya, K. Dungan, E. Hamilton, J. Hofland, S. Jan de Beur, S. Kalra, G. Kaltsas, N. Kapoor, M. Kim, C. Koch, P. Kopp, M. Korbonits, C. S. Kovacs, W. Kuohung, B. Laferrère, M. Levy, E. A. McGee, R. McLachlan, R. Muzumdar, J. Purnell, R. Rey, R. Sahay, A. S. Shah, M. A. Sperling, C. A. Stratakis, D. L. Trence, and D. P. Wilson, Eds. South Dartmouth (MA): MDText.com, Inc., 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK278967/

[12] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, X. Liu, L. Gao, T. Wu, J. Xiao, F. Wang, B. Yin, Y. Wang, G. Danala, L. He, Y. H. Choi, Y. C. Lee, S.-H. Jung, Z. Li, X. Sui, J. Wu, X. Li, T. Zhou, J. Toth, A. Baran, A. Kori, S. S. Chennamsetty, M. Safwan, V. Alex, X. Lyu, L. Cheng, Q. Chu, P. Li, X. Ji, S. Zhang, Y. Shen, L. Dai, O. Saha, R. Sathish, T. Melo, T. Araújo, B. Harangi, B. Sheng, R. Fang, D. Sheet, A. Hajdu, Y. Zheng, A. M. Mendonça, S. Zhang, A. Campilho, B. Zheng, D. Shen, L. Giancardo, G. Quellec, and F. Mériaudeau, "IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge," *Medical Image Analysis*, vol. 59, p. 101561, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841519301033

[13] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, Oct. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519305377

[14] J. Mao, X. Ma, Y. Bi, and R. Zhang, "TJDR: A High-Quality Diabetic Retinopathy Pixel-Level Annotation Dataset," Dec. 2023, arXiv:2312.15389. [Online]. Available: http://arxiv.org/abs/2312.15389

[15] M. J. Alwazzan, M. A. Ismael, and A. N. Ahmed, "A Hybrid Algorithm to Enhance Colour Retinal Fundus Images Using a Wiener Filter and CLAHE," *Journal of Digital Imaging*, vol. 34, no. 3, pp. 750–759, Jun. 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8329119/

[16] M. Cabezas and Y. Diez, "An Analysis of Loss Functions for Heavily Imbalanced Lesion Segmentation," *Sensors*, vol. 24, no. 6, p. 1981, Mar. 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/6/1981