OPTIMIZED HARD EXUDATE DETECTION WITH SUPERVISED CONTRASTIVE LEARNING

Wei Tang [1,2], *Kangning Cui* [1,2*], *Raymond H. Chan* [1,2]

[1] Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, Hong Kong
[2] Department of Mathematics, City University of Hong Kong, Hong Kong

ABSTRACT

Diabetic retinopathy (DR) is a leading global cause of blindness. Early detection of hard exudates plays a crucial role in identifying DR, which aids in treating diabetes and preventing vision loss. However, the unique characteristics of hard exudates, ranging from their inconsistent shapes to indistinct boundaries, pose significant challenges to existing segmentation techniques. To address these issues, we present a novel supervised contrastive learning framework to optimize hard exudate segmentation. Specifically, we introduce a patch-wise density contrasting scheme to distinguish between areas with varying lesion concentrations, and therefore improve the model's proficiency in segmenting small lesions. To handle the ambiguous boundaries, we develop a discriminative edge inspection module to dynamically analyze the pixels that lie around the boundaries and accurately delineate the exudates. Upon evaluation using the IDRiD dataset and comparison with state-of-the-art frameworks, our method exhibits its effectiveness and shows potential for computer-assisted hard exudate detection. The code to replicate experiments is available at github.com/wetang7/HECL/.

***Index Terms*—** hard exudate, supervised contrastive learning, medical image segmentation, deep learning.

1. INTRODUCTION

Fundus images, captured using a specialized fundus camera, provide detailed views of the interior surface of the eye, including the retina, blood vessels, and other structures [1, 2]. A pivotal application of these images is in the detection of diabetic retinopathy (DR), a complication of diabetes that affects the eyes. Among the earliest clinical signs of DR, hard exudates are notable since they are closely associated with vision damage in the early stages of DR [3]. Addressing them promptly is vital not only to prevent vision loss but also due to the increased risk of cardiovascular diseases (CVDs) associated with DR [4]. Deep learning techniques, in particular, have excelled in medical image applications, thus aiding consistent diagnosis [5, 6, 7]. To tackle the challenges of hard exudate detection, deep learning algorithms offer a more efficient and consistent alternative to manual labeling, which is labor-intensive and prone to errors [2, 8].

Despite the efficacy of existing networks, current hard exudate segmentation methods grapple with limitations due to the fine-grained, irregular shapes and non-uniform distribution of exudates across fundus images that complicate the segmentation [8, 9]. Furthermore, these exudates frequently exhibit indistinct boundaries that are neither clear nor well-defined, leading to challenges in segmentation and potentially suboptimal results [3, 10]. Addressing these challenges, supervised contrastive learning proves effective by utilizing label information to distinctly separate different classes [11], therefore managing varying lesion densities and unclear boundaries. This motivates our framework that applies supervised contrastive learning to improve hard exudate segmentation.

Our main contributions are highlighted as follows: (1) We propose a patch-wise density contrasting scheme that contrasts regions with varying lesion concentrations, which enhances the model's ability to distinguish between lesion-dense and lesion-sparse patches. (2) We design a discriminative edge inspection module using morphological operations to precisely define and analyze lesion boundaries. (3) Extensive experiments on the IDRiD dataset demonstrate our framework's effectiveness, outperforming state-of-the-art models and performing well across various backbones, thereby confirming its robustness and adaptability.

2. RELATED WORKS

2.1. Medical Image Segmentation

Deep learning has significantly enhanced medical image segmentation that helps with diagnosis [5, 6, 12, 13, 14, 15]. U-Net stands out for its integration of both a contracting and expansive path, effectively combining structure and context [5]. UNet++ introduces nested blocks and deep supervision to improve segmentation accuracy [6]while CE-Net counters U-Net's potential spatial detail loss with dense atrous convolution and multi-kernel pooling [12]. The CogSeg network refines segmentation through improved resolution and edge de-

*Corresponding Author.

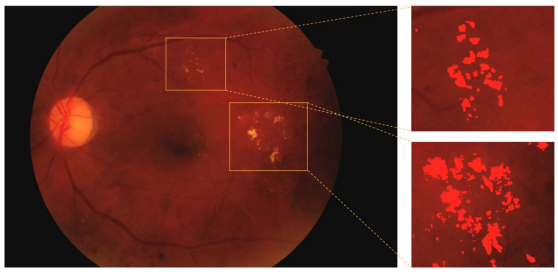

Fig. 1. Illustration of hard exudates in a fundus image. The red pixels in the right image locate the hard exudate lesions.

tection [13]. Attention U-Net emphasizes target structures using attention gates [14], and H-DenseUNet combines 2D and 3D networks for comprehensive segmentation [15]. These methods are designed for general medical image segmentation or specific applications like CT images, yet none are tailored for hard exudates with their distinct lesion structures.

2.2. Hard Exudate Detection

Hard exudate is a common clinical symptom of diabetic retinopathy, characterized by irregular white or yellowish-white accumulations in the retina resulting from plasma leakage (refer to Figure 1) [1, 2]. Manifesting as dots, patches, or circles, they serve as primary indicators of potential blindness, underscoring the critical importance of early detection and treatment [3, 8]. With the rapid advancement of deep learning techniques, several novel models have emerged for hard exudate segmentation [3, 8, 9, 10, 16, 17]. CNN-based models, in particular, utilize multi-level hierarchical information for precise segmentation [8, 9, 16, 17]. In parallel, diverse loss functions have been proposed to optimize these networks and mitigate class imbalance [9, 10, 18]. While substantial progress has been made, challenges such as the dispersion of tiny lesions and indistinct boundaries persist, necessitating continued refinement for enhanced accuracy and reliability of hard exudate detection methods.

2.3. Contrastive Learning

Contrastive learning, a self-supervised technique, learns data representations by contrasting positive (similar) and negative (dissimilar) sample pairs, mapping them closely or distantly in feature spaces [19, 20]. While unsupervised contrastive learning has shown promise in various applications, its feature representations might not be task-specific, given the absence of supervision [11]. Supervised contrastive learning incorporates labels, contrasting samples based on labels rather than mere data augmentation. This method optimizes the loss function to enhance the distinction between positive and negative pairs, yielding more discriminative features suitable for specific tasks like hard exudate detection [11, 21].

Contrastive learning, while proven for classification tasks such as DR grading, is yet to be fully explored for segmentation in DR [21]. Furthermore, the adoption of multi-level

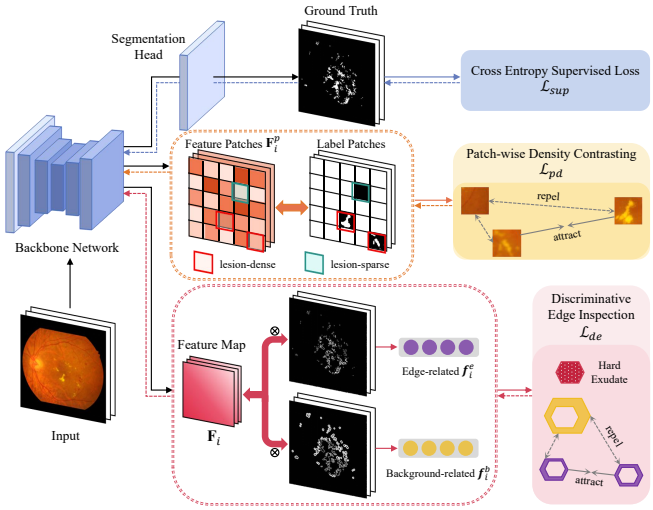

Fig. 2. An overview of the proposed framework. The network is jointly trained by $\mathcal{L}_{sup}$, $\mathcal{L}_{pd}$, and $\mathcal{L}_{pe}$ in order to learn both "density-aware" and "boundary-aware" knowledge.

feature approaches in contrastive learning shows promise in enhancing segmentation performance [22, 23]. The gap in research presents an opportunity for advancements in medical image analysis, specifically in the segmentation of hard exudates, which could benefit from the detailed feature discrimination that contrastive learning provides.

3. METHODOLOGY

Given a fundus image dataset with N samples, denoted as $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid i = 1, \ldots, N\}$, each data pair consists of an input RGB image $\boldsymbol{x}_i$ and its associated mask of hard exudates $\boldsymbol{y}_i$. Our objective is to train a model that generates a pixel-wise segmentation map $\hat{\boldsymbol{y}}_i$ for each $\boldsymbol{x}_i$, ensuring the minimization of discrepancies between $\hat{\boldsymbol{y}}_i$ and $\boldsymbol{y}_i$. Our proposed framework, depicted in Figure 2, integrates three key components: supervised training ($\mathcal{L}_{sup}$), patch-wise dense contrasting ($\mathcal{L}_{pd}$), and discriminative boundary inspection ($\mathcal{L}_{pe}$). The backbone network is refined by jointly optimizing the loss:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \alpha(\mathcal{L}_{pd} + \mathcal{L}_{de}). \tag{1}$$

The α is a hyper-parameter introduced to balance the weightings of distinct terms. The following subsections provide the motivations and detailed constructions of each component.

3.1. Patch-wise Density Contrasting

To tackle the issue of highly dispersed foreground pixel distribution and generate a more precise prediction for those fine-grained lesions, we design a patch-wise density contrasting scheme that learns from lesion pixel allocation. This approach aims to sufficiently contrast localized representations of similar and dissimilar pairs of regions, compelling the model to distinguish between the background and small clusters of hard

exudates. For each input pair $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, we divide the raw image and the label into $n \times n$ patches denoted by $\boldsymbol{x}_i^p$ and $\boldsymbol{y}_i^p$, where $p = 1, ..., n^2$. Each patch exhibits a varied distribution of hard exudate pixels, and we categorize all the $\boldsymbol{x}_i^p$'s into two sets—lesion-sparse and lesion-dense patches—by applying a proportion threshold of 0.3 for lesion pixels. Regions that differ in label distributions, i.e., lesion-sparse and lesion-dense patches, should have distinct representations in the desired feature space, thereby automatically forming negative pairs. Given a feature patch $\mathbf{F}_i^p$ of size $C \times h \times w$, we regularize it into a one-dimensional vector $\boldsymbol{f}_i^p$ of size $C \times 1$:

$$f_{i,c}^p = \frac{\sum_{j=1}^{h} \sum_{k=1}^{w} F_{i,c,j,k}^p}{h \times w}, \quad c = 1, 2, \ldots, C. \tag{2}$$

Here, $F_{i,c,j,k}^p$ is the value at channel c, height j, and width k of the feature map for patch $\boldsymbol{x}_i^p$. The sum is averaged over spatial dimensions for each channel. After obtaining $\boldsymbol{f}_i^p$'s, we exploit the supervised contrastive loss to enhance the model's capacity to distinguish small patches of hard exudates from the background. The patch-wise density contrastive loss $\mathcal{L}_{pd}$ for an entire mini-batch is formulated as:

$$\mathcal{L}_{pd} = \frac{1}{|\mathcal{M}|} \sum_{\boldsymbol{f}_i^p \in \mathcal{M}} \mathcal{L}(\boldsymbol{f}_i^p, \mathcal{P}(p), \mathcal{N}(p)), \tag{3}$$

with $\mathcal{L}(\boldsymbol{f}_i^p, \mathcal{P}(p), \mathcal{N}(p)) =$

$$\frac{-1}{|\mathcal{P}(p)|} \sum_{\boldsymbol{f}_i^q \in \mathcal{P}(p)} \log \frac{\exp(\mathrm{sim}(\boldsymbol{f}_{\mathrm{i}}^{\mathrm{q}}, \boldsymbol{f}_{\mathrm{i}}^{\mathrm{p}})/\tau)}{\sum_{\boldsymbol{f}_{\mathrm{i}}^{\mathrm{k}} \in \mathcal{P}(\mathrm{p}) \cup \mathcal{N}(\mathrm{p})} \exp(\mathrm{sim}(\boldsymbol{f}_{\mathrm{i}}^{\mathrm{k}}, \boldsymbol{f}_{\mathrm{i}}^{\mathrm{p}})/\tau)}, \tag{4}$$

where $|\cdot|$ counts the cardinality of a set, $\mathcal{P}(p)$ denotes the set of positives that excludes $\boldsymbol{f}_i^p$, $\mathcal{N}(p)$ is the corresponding negative set, and $\mathcal{M}$ represents the set of stored representative features in every mini-batch. The cosine similarity function $\mathrm{sim}(\cdot, \cdot)$ is applied here to measure the similarity between two feature vectors, and τ stands for the temperature.

3.2. Discriminative Edge Inspection

Failure to identify the blurry boundaries and differentiate them from the surrounding tissue is another contributor to the degradation of segmentation performance. To address this, we introduce a discriminative edge inspection module to dynamically analyze the pixels around the lesion boundaries and then correctly separate them into the corresponding groups. For given patches $(\boldsymbol{x}_i^p, \boldsymbol{y}_i^p)$, morphological operations are utilized to extract the inner and outer contour masks of each patch's binary ground truth $\boldsymbol{y}_i^p$. The inner contour, containing only lesion pixels within the edge, is derived by subtracting the eroded label map from the original mask $\boldsymbol{y}_i^p$. The outer contour is obtained by subtracting $\boldsymbol{y}_i^p$ from the dilated one. Specifically, for lesion-dense patches, we set the iterations for dilation and erosion to two. For lesion-sparse patches, which typically have scattered pathological spots and a vast background, the iterations are set to five for dilation and one for erosion to gain extra information about the background. We compose the patches together to form the final inner contour $\boldsymbol{y}_{in}$ and outer contour $\boldsymbol{y}_{out}$ after the morphological operations, and we compute the corresponding averaged edge-related feature $\boldsymbol{f}_i^e$ and background-related feature $\boldsymbol{f}_i^b$ inspired by [24]:

$$\boldsymbol{f}_i^e = \frac{\sum_{j=1}^{H} \sum_{k=1}^{W} (\mathbf{F}_i \otimes \boldsymbol{y}_{in})_{jk}}{\sum_{j=1}^{H} \sum_{k=1}^{W} (\boldsymbol{y}_{in})_{jk}}, \quad \boldsymbol{f}_i^b = \frac{\sum_{j=1}^{H} \sum_{k=1}^{W} (\mathbf{F}_i \otimes \boldsymbol{y}_{out})_{jk}}{\sum_{j=1}^{H} \sum_{k=1}^{W} (\boldsymbol{y}_{out})_{jk}} \tag{5}$$

where $\otimes$ represents element-wise multiplication, $\mathbf{F}_i$ is the corresponding feature map, H and W are the height and width. Finally, the loss $\mathcal{L}_{de}$ is defined as:

$$\mathcal{L}_{de} = \frac{1}{2|\mathcal{B}|} \sum_{\boldsymbol{f}^t \in \{\boldsymbol{f}_i^e, \boldsymbol{f}_i^b\}} \mathcal{L}(\boldsymbol{f}^t, \mathcal{P}(t), \mathcal{N}(t)). \tag{6}$$

Here, $\mathcal{B}$ denotes the images in each batch. If $\boldsymbol{f}^t$ is an edge-related feature, then $\mathcal{P}(t)$ is the set of other edge-related features, and $\mathcal{N}(t)$ designates all the background-related features in a mini-batch, and vice versa. With this module, the model is expected to learn more informative representations of the pixels around the lesion boundary, and therefore develop the necessary ability to recognize the ambiguous edges.

4. EXPERIMENTS

We evaluate our proposed method on a benchmark dataset: the Indian diabetic retinopathy image dataset (IDRiD) [25] due to its high-quality annotations. Specifically, we use the subset annotated for hard exudates. This subset comprises 81 color fundus JPEG images of resolution 4288×2848, with 54 officially designated for training. Given the dataset's limited size, we employ data augmentation techniques such as random horizontal flipping, rotation (ranging from -180° to +180°), and adjustments in brightness and contrast (scaling between 50% and 150%) to mitigate overfitting. For model input, all images are resized and cropped to 512×512 pixels.

4.1. Experimental Setup

We employ UNet++ [6] as our base network due to its prevalence in medical image segmentation tasks. The model is optimized using Adam optimizer with a batch size of 6. The initial learning rate is set as 3×10^{-4} and decays by 0.5 after every 100 epochs. For the hyper-parameters, images are divided into 5×5 patches, and the temperature parameter τ is set as 0.05. The balancing parameter α is empirically set as 0.01. To guarantee model convergence, training persists for 300 epochs for all models. The experiments are executed on an NVIDIA GeForce RTX 3090 GPU using PyTorch [26].

Model	IoU	F_1 score	AUC	Recall
H-DenseUNet	44.59	59.71	**98.50**	52.13
Att-UNet	51.81	66.75	95.67	59.66
DeepLabv3+	41.10	58.21	84.07	47.81
HED	52.68	<u>68.96</u>	38.98	<u>65.50</u>
CogSeg	46.63	63.44	63.36	55.56
Ours	**55.69**	**69.64**	96.59	**65.65**

Table 1. Comparative performance of our method with state-of-the-art networks on hard exudates segmentation. Bold values indicate the best performance for each metric, while underlined values denote the second-best. For a fair comparison, we list the best performance of each model.

Method	IoU	F_1 score	AUC	Recall
U-Net	50.61 (0.26)	65.48 (0.26)	**96.81** (0.24)	59.11 (0.58)
U-Net w/ Proposed	**52.10** (0.21)	**66.96** (0.25)	95.87 (0.19)	**61.08** (0.60)
ResUnet	38.70 (3.71)	54.80 (3.51)	95.58 (0.94)	42.12 (3.84)
ResUnet w/ Proposed	**41.09** (1.39)	**56.57** (1.56)	**96.81** (0.71)	**46.80** (1.90)
CE-Net	37.67 (2.57)	52.33 (2.89)	95.26 (0.60)	38.51 (3.34)
CE-Net w/ Proposed	**46.26** (0.86)	**61.76** (0.89)	**96.33** (1.47)	**51.58** (2.68)
UNet++	53.30 (0.65)	67.78 (0.41)	**97.13** (0.79)	63.35 (3.42)
UNet++ w/ $\mathcal{L}_{pd}$	54.24 (0.19)	68.38 (0.20)	<u>97.00</u> (1.08)	<u>63.93</u> (1.19)
UNet++ w/ $\mathcal{L}_{de}$	<u>54.91</u> (0.36)	<u>68.85</u> (0.36)	96.32 (0.51)	63.74 (0.77)
UNet++ w/ Proposed	**55.38** (0.25)	**69.34** (0.24)	96.50 (0.08)	**64.86** (0.74)

Table 2. Average results of the proposed modules combined with other CNN backbones for hard exudate segmentation. Numbers in parentheses are standard deviations.

For comparative analysis, we compare our approach against state-of-the-art networks including UNet++[6], H-DenseUNet[15], and Att-UNet [14] that we implemented, as well as DeepLabv3+[27], HED[18], and CogSeg [13], whose results we directly adopt from [17]. Additionally, extensive ablation studies are conducted, assessing the impact of different components in the proposed loss and the robustness adaptability of our framework with alternative backbones like U-Net [5], ResUnet [28], and CE-Net [12]. We repeat the experiment 5 times to obtain the mean and standard deviation.

4.2. Results and Discussion

The comparative results of the proposed framework and other networks are shown in Table 1, measured in four metrics: IoU, F_1 score, AUC, and recall. Our framework performs the best in IoU, F_1 score, and recall. H-DenseUNet, despite its exceptional AUC, falters in other metrics; meanwhile, HED, with a promising recall and F_1 score, struggles due to a notably low AUC, both suggesting insufficient segmentation. Moreover, the inclusion of either $\mathcal{L}_{pd}$ or $\mathcal{L}_{de}$ enhances certain metrics relative to the UNet++ backbone, as shown in Table 2. Yet, the fusion of both $\mathcal{L}_{pd}$ and $\mathcal{L}_{de}$ in our proposed approach yields the best results with only a marginal AUC trade-off.

Table 2 highlights the enhanced performance of the proposed framework over the vanilla networks, evidenced across three different backbones. The proposed method delivers substantial improvements in all evaluated metrics, indicating a segmentation that is both more precise and reliable. The qualitative evidence, shown in Figure 3, further confirms this quantitative assessment. Our method proficiently identifies subtle lesions and ambiguous boundaries, while successfully preventing the misclassification of structures such as the optic disc. Furthermore, our ablation study demonstrates the generalizability of the proposed framework that consistently enhances the performance of various network architectures.

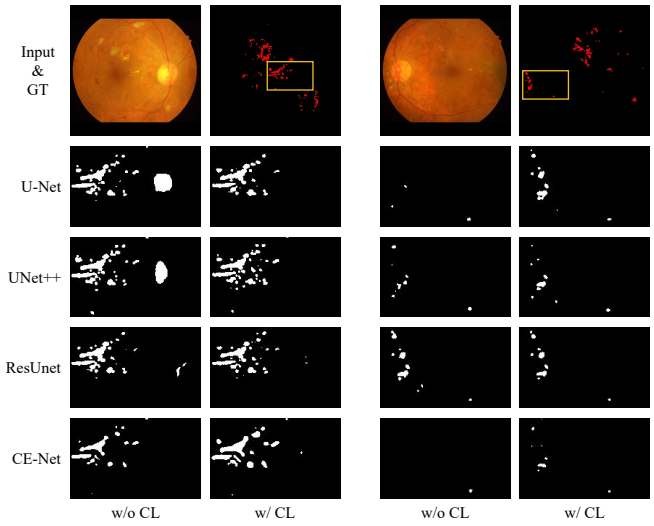

Fig. 3. Comparison of network segmentations with (w/ CL) and without (w/o CL) the proposed framework. Top row: input images and GT masks. Following rows: segmentation outputs, with yellow boxes emphasizing optic disc and small lesion areas.

5. CONCLUSION AND FUTURE WORKS

The study has shown that integrating the patch-wise density contrasting scheme and discriminative edge inspection module by supervised contrastive losses considerably improves the accuracy of hard exudate detection. The robustness and adaptability of our framework have been further confirmed through comprehensive ablation studies. Future work will extend our framework for segmentation across additional hard exudate datasets with domain shift. Given its generalizability, our proposed method may prove effective for various medical imaging tasks that present challenges similar to those of hard exudate detection.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by [25]. Ethical approval was not required as confirmed by the license attached with the open access data.

7. ACKNOWLEDGEMENT

This work was partially supported by HKRGC GRF grants CityU1101120, CityU11309922, CRF grant C1013-21GF, and HKRGC-NSFC Grant N_CityU214/19. The authors would like to thank Dr. Jizhou Li and the Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (hkcoche.org) for the collaboration and support in this research.

8. REFERENCES

[1] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Inform. Med. Unlocked*, vol. 20, pp. 100377, 2020.

[2] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artif. Intell. Med.*, vol. 99, pp. 101701, 2019.

[3] C. I. Sánchez, M. García, A. Mayo, M. I. López, and R. Hornero, "Retinal image analysis based on mixture models to detect hard exudates," *Med. Image Anal.*, vol. 13, no. 4, pp. 650–658, 2009.

[4] N. Cheung and T. Y. Wong, "Diabetic retinopathy and systemic vascular complications," *Prog. Retin. Eye Res.*, vol. 27, no. 2, pp. 161–176, 2008.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc MICCAI*. Springer, 2015, pp. 234–241.

[6] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support.* Springer, 2018, pp. 3–11.

[7] F. Pan et al., "Dual-view selective instance segmentation network for unstained live adherent cells in differential interference contrast images," *arXiv preprint arXiv:2301.11499*, 2023.

[8] Z. Si, D. Fu, Y. Liu, and Z. Huang, "Hard exudate segmentation in retinal image with attention mechanism," *IET Image Process.*, vol. 15, no. 3, pp. 587–597, 2021.

[9] J. Mo, L. Zhang, and Y. Feng, "Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks," *Neurocomputing*, vol. 290, pp. 161–171, 2018.

[10] S. Guo et al., "Bin loss for hard exudates segmentation in fundus images," *Neurocomputing*, vol. 392, pp. 314–324, 2020.

[11] P. Khosla et al., "Supervised contrastive learning," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 18661–18673, 2020.

[12] Z. Gu et al., "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[13] Y. Sang, J. Sun, S. Wang, H. Qi, and K. Li, "Super-resolution and infection edge detection co-guided learning for covid-19 ct segmentation," in *Proc. ICASSP*. IEEE, 2021, pp. 1665–1669.

[14] O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[15] X. Li et al., "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[16] S. Guo et al., "L-seg: An end-to-end unified framework for multi-lesion segmentation of fundus images," *Neurocomputing*, vol. 349, pp. 52–63, 2019.

[17] J. Zhang et al., "Hard exudate segmentation supplemented by super-resolution with multi-scale attention fusion module," in *Proc. BIBM*. IEEE, 2022, pp. 1375–1380.

[18] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. ICCV*, 2015, pp. 1395–1403.

[19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. ICML*. PMLR, 2020, pp. 1597–1607.

[20] S. Camalan et al., "Detecting change due to alluvial gold mining in peruvian rainforest using recursive convolutional neural networks and contrastive learning," in *Fall Meeting*. AGU, 2022.

[21] M. R. Islam et al., "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," *Comput. Biol. Med.*, vol. 146, pp. 105602, 2022.

[22] H. Wu, Z. Wang, Y. Song, L. Yang, and J. Qin, "Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images," in *Proc. CVPR*, 2022, pp. 11666–11675.

[23] X. Zhao et al., "Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation," in *Proc. ISBI*. IEEE, 2022, pp. 1–5.

[24] Q. Liu, C. Chen, J. Qin, Q. Dou, and P. Heng, "Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proc. CVPR*, 2021, pp. 1013–1023.

[25] P. Porwal et al., "Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, pp. 25, 2018.

[26] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[27] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, 2018, pp. 801–818.

[28] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018.