

ReSynthDetect: A Fundus Anomaly Detection Network with Reconstruction and Synthetic Features

Jingqi Niu¹

stu_jqniu@sjtu.edu.cn

Qinji Yu¹

cscyqj@sjtu.edu.cn

Shiwen Dong¹

1481005290@sjtu.edu.cn

Zilong Wang²

zlwang@voxelcloud.net.cn

Kang Dang² †

kdang@voxelcloud.net.cn

Xiaowei Ding^{1,2} †

dingxiaowei@sjtu.edu.cn

¹ Department of Electronic Engineering

Shanghai Jiao Tong University

Shanghai, China

² Voxelcloud Inc.

Chenhang Road, Minhang District,

Shanghai, China

Abstract

Detecting anomalies in fundus images through unsupervised methods is a challenging task due to the similarity between normal and abnormal tissues, as well as their indistinct boundaries. The current methods have limitations in accurately detecting subtle anomalies while avoiding false positives. To address these challenges, we propose the ReSynthDetect network which utilizes a reconstruction network for modeling normal images, and an anomaly generator that produces synthetic anomalies consistent with the appearance of fundus images. By combining the features of consistent anomaly generation and image reconstruction, our method is suited for detecting fundus abnormalities. The proposed approach has been extensively tested on benchmark datasets such as EyeQ and IDRiD, demonstrating state-of-the-art performance in both image-level and pixel-level anomaly detection. Our experiments indicate a substantial 9% improvement in AUROC on EyeQ and a significant 17.1% improvement in AUPR on IDRiD.

1 Introduction

Deep Convolutional Neural Networks (CNNs) have made significant breakthroughs in various relevant fields of medical image analysis [10, 25, 30]. However, current fully supervised methods require a vast amount of annotated abnormal images. Obtaining such images can be challenging, particularly for rare diseases with low incidence rates. Conversely, collecting normal images is relatively easier. Therefore, recent research has focused on unsupervised anomaly detection methods [9, 15, 23] that identify anomalies in medical images

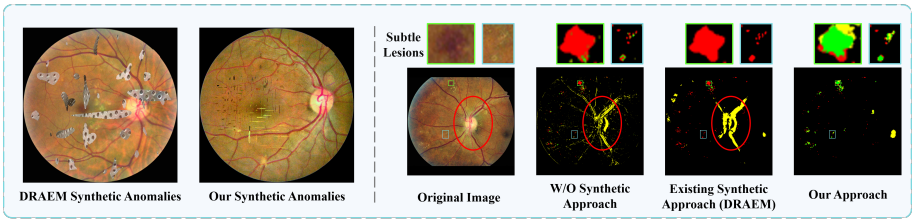


Figure 1: Our method generates more consistent lesions compared to existing methods (in left figure). Therefore, the existing methods have false positives in normal structures (see red circles), and our method has better detection performance (green area = true positives, yellow area = false positives, red area = missing).

through training only on normal images. Among medical imaging modalities, fundus image anomaly detection presents an especially challenging scenario. Retinal lesions come in various shapes, sizes, and textures. Learning them in an unsupervised manner can be difficult, particularly when they have indistinct boundaries and are visually similar to normal fundus tissues. Consequently, the development of an accurate and reliable method for detecting anomalies in fundus images is a critical area of ongoing research.

Currently, most anomaly detection methods rely on either reconstruction or representation based approaches [14, 16, 27]. However, these methods are typically trained solely on non-anomalous data and may not be optimized for discriminative anomaly detection. Consequently, they face challenges in learning abnormality representations and distinguishing lesions that are not significantly different from normal tissues. This can lead to inaccurate detection of subtle lesions in fundus images and falsely identifying normal areas as anomalies, as shown in Fig. 1. To address this issue, recent studies [12, 28] have employed synthetic anomalies. Nevertheless, these techniques, such as DRAEM [26], may produce inconsistent anomalies that do not match the appearance of fundus images, which can be misleading. To overcome these limitations, we propose a novel approach that includes an anomaly generator capable of producing anomalies consistent with the appearance of fundus images. Additionally, to prevent overfitting to synthetic anomalies, we have implemented a reconstruction network which effectively reconstructs and models normal images. The reconstructive features produced by this network are combined with synthetic anomaly features to accurately localize any anomalies. As demonstrated in Fig. 1, our approach has been successful in detecting subtle fundus lesions while minimizing false positives by avoiding the misidentification of normal structures as anomalies.

Our proposed approach, named ReSynthDetect network which combines reconstruction and synthetic features, is designed for detecting fundus anomalies. The network is trained in two stages as shown in Fig. 2. In the first stage, a reconstruction network is trained on normal images, while in the second stage, an anomaly localization network is trained using artificially created anomalies. By incorporating information from the reconstruction network, the localization network accurately identifies anomalies while reducing reliance on synthetic anomalies. We create synthetic lesions by randomly selecting normal training images as source and target images. We augment the source images to create diverse lesions, which are then pasted onto random positions on the target images using self-mix [5]. This approach ensures consistency in the fundus images while simulating variations in real retinal lesions. Furthermore, we conducted comprehensive experiments on network architecture and found that, unlike previous literature [26], the best results for retina anomaly detection

were obtained by combining the encoder features of the reconstruction network and the encoder features of the anomaly localization network. Through careful anomaly generation and network architecture selection, our proposed approach achieved state-of-the-art results on both the EyeQ [24] and IDRiD [18] benchmark datasets.

Contributions. (1) We propose a new approach named ReSynthDetect network designed for detecting anomalies in fundus images by combining reconstruction and synthetic features. (2) We introduce a novel anomaly generator that can produce diverse and consistent synthetic anomalies in fundus images. (3) Our proposed approach achieves state-of-the-art results on two benchmark retinal datasets, EyeQ and IDRiD, with a 9% improvement in AU-ROC for image-level anomaly detection on EyeQ and a 17.1% improvement in AUPR for pixel-level anomaly localization on IDRiD.

2 Related Work

Most of the existing anomaly detection methods are reconstruction-based [2, 8, 29], which train models to reconstruct normal data during training and detect anomalies by calculating the reconstruction error. fAnoGAN in [23] applies an adversarial network for normal image reconstruction and calculates anomaly scores by reconstruction error. WDMT in [27] use a weight-decay skip connection strategy for reconstruction network and integrating an auxiliary task of the histogram of oriented gradients prediction to improve feature representation learning. Lesion2Void in [9] masks out normal patches and trains a reconstruction model based on the correlation with neighboring pixels to distinguish anomalies. However, these methods often have relatively large reconstruction errors in normal retinal structures such as optic disc, cup, and blood vessels, resulting in potential false positives [24].

To solve this problem, some recent representation-based methods have been proposed [9, 26, 20], which compute anomaly scores based on the similarity of the features between the test and normal samples. ReSAD in [15] extracts features by a pre-trained model and proposes a spatial and region module for local and long-range anomaly detection. MKD in [22] applies knowledge distillation between a pre-trained source network and a smaller cloner network and calculates feature similarity as anomaly scores. Nevertheless, representation-based methods lack the reference of abnormal features, making it challenging to detect subtle lesions in fundus images.

A few works attempt to utilize synthetic anomalies for anomaly detection [12, 24, 28]. DRAEM [26] trains a reconstructive network on synthetic anomalies and utilizes a discriminative network to detect deviations from synthetic and reconstructed images as anomalies. However, the reconstructed images may contain deviations in normal structures, which can be falsely detected as anomalies. Our work is built on a similar architecture as described in [26], but we introduce two key technical differences to overcome its limitations. Firstly, we propose a novel anomaly generator that can produce diverse and consistent synthetic anomalies in fundus images. Secondly, unlike previous literature [26], we achieve the best results for retina anomaly detection by combining the encoder features of both the reconstruction network and the anomaly localization network.

3 Method

The pipeline of our proposed method is illustrated in Fig. 2. In the first stage, a reconstruction network is trained on normal training images (Sec. 3.1), and its encoder is utilized

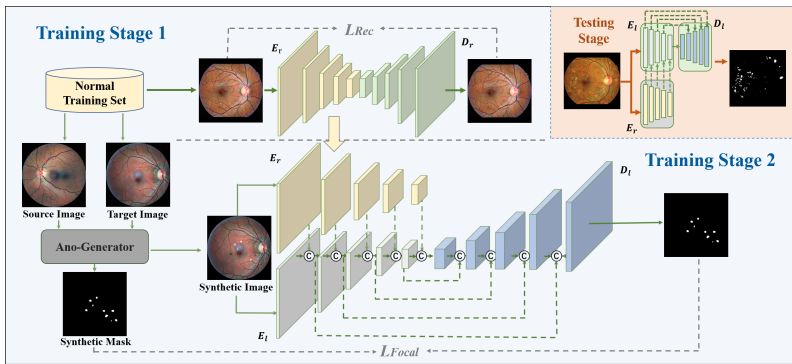


Figure 2: **Overall Pipeline:** In the first training stage, we train a reconstruction network to reconstruct normal images in order to obtain its encoder E_r as a feature extractor. In the second stage, we concatenate the reconstructive feature extract from E_r to a U-shape localization network for synthetic anomalies localization. Finally, we test our model in the testing set by utilizing the localization network with reconstructive features extracted by E_r .

as a feature extractor in the subsequent training stage. In the second stage, a reconstructive feature-guided anomaly localization network is trained using synthetic anomalies (Sec. 3.2). The synthetic anomalies are obtained by a consistent anomaly generator (Sec. 3.3).

3.1 Reconstruction Based Feature Extractor

Relying solely on synthetic anomalies can result in overfitting to their specific patterns. Previous work, DRAEM [26], combines reconstructed images with synthetic anomalies to identify deviations as anomalies. However, reconstructed images may contain deviations in normal structures, leading to false positives. To overcome this, we train a reconstruction network as a feature extractor, which mitigates overfitting to synthetic anomaly patterns and avoids false positives in normal structures.

In the first training stage, an autoencoder is utilized as the reconstruction network, with the objective of reconstructing the normal fundus images which serve as the input of the network. This process enables the extraction of reconstructive features, which are subsequently utilized in the second stage of training

Formally, the reconstruction network comprised of an encoder E_r and a decoder D_r , is trained on the normal training fundus image I . We utilize the L2 loss as the reconstruction loss, which can be calculated as follow:

$$L_{Rec} = \|D_r(E_r(I)) - I\|_2^2. \quad (1)$$

Once the training of the autoencoder is completed, its parameters are fixed and will no longer be changed in the following stage. The encoder E_r of the reconstruction network will be used as the extractor of the reconstructive feature, while the decoder D_r will be discarded.

3.2 Reconstructive Features Guided Localization Network

Due to the lack of real anomaly samples in the training phase, the network needs to be trained on proxy tasks. We utilize the localization of synthetic anomalies as the proxy task to train the network.

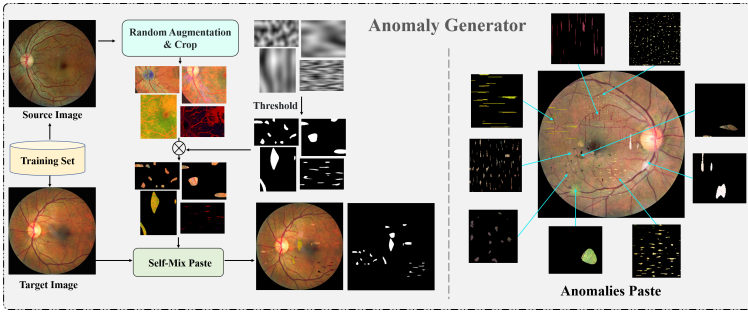


Figure 3: **Anomaly Generator**: The left image shows the process of the anomaly generator, which creates lesions from a source image and pastes the generated lesions onto the target image. The image on the right demonstrates that the proposed method is capable of generating lesions with various shapes and textures. These lesions can be pasted onto random locations of the target images without disrupting the continuity of the image. More synthetic images can be found in the Supplementary.

In the second training stage, a reconstructive feature-guided localization network is trained on the synthetic image I_G and its corresponding mask M_G (see Sec. 3.3). We apply a U-shape model [14] with skip connections as the localization network, which consists of an encoder E_l and a decoder D_l , to localize the synthetic anomalies. We extract both reconstructive features extracted by E_r and localization features extracted by E_l in the synthetic image I_G and concatenate them in each layer.

More specifically, at each layer i of the encoders E_r and E_l , we extract the corresponding reconstructive features F_r^i and localization features F_l^i , respectively, and F_c^i is obtained by concatenating them as $F_c^i = \text{concat}(F_r^i, F_l^i)$. Subsequently, F_c^i is used as input in the subsequent layers of encoder E_l and decoder D_l for the localization task.

The Focal Loss [15] is introduced during training to alleviate the imbalance between the normal pixels and anomalous pixels. It is expressed as follows:

$$L_{Focal} = \begin{cases} -(1-p)^\tau \log(p), & M_G^{x,y} = 1, \\ -p^\tau \log(1-p), & M_G^{x,y} = 0. \end{cases} \quad (2)$$

Here, p represents the probability of anomaly at position (x, y) predicted by the model, and τ is a tunable focusing parameter, and set to 2 in this paper.

3.3 Consistent Anomaly Generator

As depicted in Fig. 3, our approach incorporates an anomaly generator that generates lesions based on the source image and subsequently pastes the generated lesions onto the target image. In order to maintain the consistent appearance between normal retinal images and synthetic anomalies, we randomly sample a source image I_s and a target image I_t from the normal training set. To generate a variety of texture anomalies, we randomly select three augmentation methods from our pool of candidates, which include sharpening, solarizing, gamma contrast enhancement, hue change, color temperature alteration, auto-contrast, and random color shifting. We apply these augmentations to the source image I_s and produce an augmented image. Subsequently, the anomaly generator performs a random cut to obtain crop C_s of variable size from the augmented source image at a random location to generate lesions.

To generate anomalies with diverse shapes, we use Perlin noise, a type of gradient noise commonly employed in computer graphics. Our anomaly generator utilizes a Perlin noise generator [14] to produce Perlin noise \mathbf{P}_n of the same size as \mathbf{C}_s . Subsequently, a thresholding process is applied to generate a binary mask \mathbf{P} from \mathbf{P}_n .

Directly pasting the augmented source crop \mathbf{C}_s onto target images can potentially introduce inconsistencies in the boundaries of the pasted lesions. To address this issue, we use self-mix paste module, which uses Euclidean Distance Transform algorithm [6] to compute the distance between each pixel in the mask \mathbf{P} and its nearest background pixel, generating a distance map \mathbf{D} . Subsequently, fusion weights map \mathbf{W} are generated according to the Eq. 3:

$$\mathbf{W} = (1 - \alpha) \times \frac{\mathbf{D} - \min(\mathbf{D})}{\max(\mathbf{D}) - \min(\mathbf{D})} + \alpha. \quad (3)$$

where α is the scaling factor, and set to 0.7 in this paper.

Next, self-mix paste module selects a crop \mathbf{C}_t in the target image with a random location and fuse the \mathbf{C}_s and \mathbf{C}_t with \mathbf{W} and \mathbf{P} according to Eq. 4. Notably, \mathbf{C}_s and \mathbf{C}_t have the same size but may not be located at the same position.

$$\begin{aligned} \mathbf{C} &= (\mathbf{P} \odot \mathbf{W}) \odot \mathbf{C}_s + (1 - \mathbf{P} \odot \mathbf{W}) \odot \mathbf{C}_t, \\ \mathbf{C}_m &= \mathbf{P}, \end{aligned} \quad (4)$$

where \odot denotes the element-wise product, $\mathbf{P} \odot \mathbf{W}$ denote a smoothing mask.

In Eq. 4, \mathbf{C} denotes the generated anomaly crop, the \mathbf{C}_m denotes the corresponding mask. Finally, the anomaly generator can obtain the synthetic image \mathbf{I}_G and corresponding mask \mathbf{M}_G through \mathbf{C} and \mathbf{C}_m , which is utilized in previous training.

4 Experiments

4.1 Experimental Protocol

Datasets. For evaluation, we used two public datasets: EyeQ [15] for image-level anomaly detection and IDRiD [16] for pixel-level anomaly localization. We applied Contrast Limited Adaptive Histogram Equalization (CLAHE) [17] with a ClipLimit of 2 and a GridSize of 8 to enhance image contrast while preserving local details. The input image size for all datasets was standardized to 768×768 for consistency.

- **EyeQ:** The EyeQ [15] dataset is a subset of the EyePACS [5] dataset used for grading diabetic retinopathy (DR). EyeQ consists of 28,792 fundus photographs with DR grading annotations and corresponding image quality labels. The images in the EyeQ dataset are classified into "good", "usable", or "reject" categories based on the image quality, and only the images classified as "good" are used in our experiments. The DR disease severity in the EyeQ dataset is divided into five grades: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), and 4 (proliferative) [18]. Following [9], images with level 0 are considered as normal images, and images with grades 1 - 4 are considered as abnormal images. During the training phase, we utilizes all the normal images in training set from the EyeQ dataset, which consist of 6,324 normal images. For testing, we used the complete testing set comprising 8,470 images from the EyeQ dataset.
- **IDRiD:** We used the Indian Diabetic Retinopathy Image Dataset (IDRiD) [16] dataset, which contains highly precise DR lesion masks and is commonly used as a benchmark

dataset for lesion localization tasks. Specifically, we used 134 normal retinal images as the training set, 32 normal retinal images and 81 abnormal retinal images with finely annotated DR lesions as the testing set. The abnormal image contains four types of retinal lesions with fine annotated masks, including microaneurysms (MA), soft exudates (SE), hard exudates (EX), and hemorrhages (HE).

Implementation Details. The codes are implemented using PyTorch on a single NVIDIA RTX 3090 GPU with 24GB memory. The initial learning rate is set to $5e-5$ with a cosine learning rate decay, reaching a minimum learning rate of $2.5e-5$. Additionally, a warm-up strategy with a duration of 50 epochs is implemented. For image-level anomaly detection, we compute the anomaly score by averaging the highest predicted anomaly probabilities of the top 10 pixels in the model’s output.

Evaluation Metric. We evaluate the performance using the standard metric for anomaly detection, AUROC, for both image-level anomaly detection and pixel-level anomaly localization. However, the AUROC can not precisely reflect the localization result especially in fundus images anomaly detection. The reason is that false positive rate is dominated by the *a-priori* very high number of non-anomalous pixels and is thus kept low despite of false positive detections. We thus additionally report the pixel-wise Area Under the Precision-Recall curve (AUPR), which is more realistic for the lesion localization performance [14, 21], especially for retinal lesion localization performance [18] because it is more appropriate for highly imbalanced classes. Besides, we also evaluate the performance of the method by balanced accuracy (ACC) on pixel-level anomaly detection in order to partially mitigate the issue of imbalanced distribution of positive and negative samples.

Method	0vs1	0vs2	0vs3	0vs4	0 vs all
fAnoGAN [23]	0.508	0.491	0.525	0.577	0.514
MKD [22]	0.582	0.547	0.623	0.706	0.546
DRAEM [26]	0.585	0.658	0.742	0.716	0.614
Lesion2Void [9]	0.567	0.625	0.877	0.902	0.632
Ours	0.556	0.764	0.941	0.919	0.722

Table 1: Image Level Anomaly Detection Result for EyeQ in AUROC.

Method	AUROC	ACC	AUPR
fAnoGAN [23]	0.756	0.686	0.048
MemAE [8]	0.749	0.596	0.058
WNet [27]	0.775	0.566	0.075
DRAEM [26]	0.827	0.747	0.100
ReSAD [15]	0.905	0.819	0.256
Ours	0.931	0.859	0.427

Table 2: Pixel Level Anomaly Detection Result for IDRiD.

4.2 Comparisons with the State of the Arts

Image-level Anomaly Detection. Following [9], we compare our proposed method with multiple SOTAs: two reconstruction-based methods: fAnoGAN [23] and Lesion2Void [9], a synthetic anomalies based method DRAEM [26], a representation-based method MKD [22] as the baseline model for image-level anomaly detection.

Tab. 1 quantitatively compares our model with baselines in the image-level anomaly detection on EyeQ. Grade 0 is considered as a normal image. For the comparisons of 0 vs 1, 0 vs 2, ..., 0 vs 4, we consider only DR graded images from grade 1 to grade 4 as abnormal images. For the comparisons of 0 vs all, we use all abnormal images with DR grades from 1 to 4 for anomaly detection. Our approach surpasses all the baselines in 0 vs all grade experiments and surpasses the previous best SOTA method by 9%, which demonstrates that our method can achieve state-of-the-art performance in image-level retinal anomaly detection.

Pixel-level Anomaly Localization. We compare our method with multiple SOTA methods: three reconstruction-based methods: fAnoGAN [23], MemAE [8] and WDMT-Net [22], a synthetic anomalies based method DRAEM [26] and a representation-based approach ReSAD [15] as our baseline models.

Tab. 2 quantitatively compares our model with recent approaches on the pixel-level anomaly localization. Compared to the best state-of-the-art methods, our approach shows improvements of 2.6% in AUROC, 4% in ACC, and a significant increase of 17.1% in AUPR. This suggests that our method achieves precise anomaly localization results for retinal lesions (which is supported by Fig. 4). In summary, our approach has achieved the best performance in the pixel-level anomaly localization task for retinal lesions.

Ablation Method					Metric			
Loc.Net.	Rec.Net.		Concatenate Type		IDRiD			EyeQ(0vsall)
	Random	Train	Image	Feature	AUROC	ACC	AUPR	AUROC
		✓			0.801	0.719	0.083	0.485
✓					0.891	0.815	0.331	0.656
✓	✓			✓	0.883	0.817	0.267	0.682
✓		✓	✓		0.819	0.742	0.149	0.581
✓		✓		✓	0.931	0.859	0.427	0.722

Table 3: Ablation on Architecture and Concatenate Type (IDRiD for pixel-level and EyeQ for image-level): "Loc.Net." indicates the use of a localization network, "Rec.Net." indicates the use of a reconstruction network ("Random" for random initialization and "Train" for a trained reconstruction network on the proxy task). "Image" concatenates reconstructed images (similar to DRAEM), while "Feature" concatenates reconstructive features (like our proposed method).

Ablation Method			Metric			
Source Image	Self-Mix	Mask \mathbf{P}	IDRiD			EyeQ (0vsall)
			AUROC	ACC	AUPR	AUROC
Texture	✓	✓	0.849	0.747	0.181	0.666
Fundus	✓		0.781	0.720	0.043	0.560
Fundus		✓	0.890	0.819	0.274	0.682
Fundus	✓	✓	0.931	0.859	0.427	0.722

Table 4: Ablation on Anomaly Generation (IDRiD for pixel-level and EyeQ for image-level).

4.3 Ablation Study

Influences of network architectures. Table 3 presents the results of ablation experiments on architecture and concatenate type. Without the localization network (first line), utilizing only the autoencoder in the first training stage and calculating reconstruction error as the anomaly

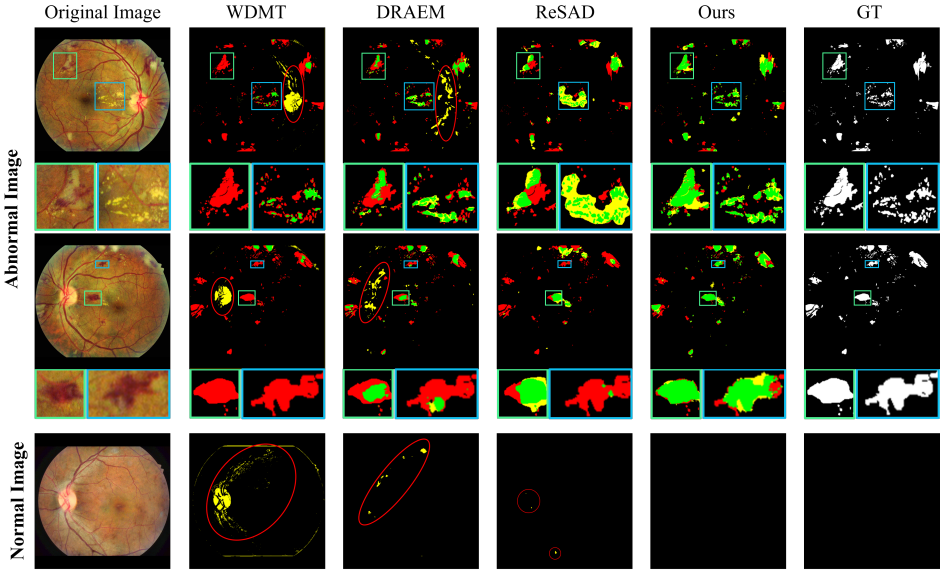


Figure 4: **Lesion Localization Results:** The green area in the figure represents true positives, the yellow area represents false positives, and the red area represents missing detection. Red circles denote the false positive in the normal structure of fundus

score leads to a significant performance drop. This underscores the importance of introducing the localization network for synthetic anomaly localization. In the ablation experiment of the reconstruction network (second and third lines), the absence of the reconstruction network or using randomly initialized network leads to a notable performance drop, emphasizing the significance of the reconstruction network. Additionally, concatenating restored images (fourth line) results in a significant performance degradation, indicating that image concatenation may not be suitable for fundus anomaly detection. Comparing the results in the second line with the third and fourth lines, we observe that concatenating random features or restored images leads to a reduction in model performance, indicating the presence of misleading information will misguide the model and decrease its overall performance.

Influences of anomaly generation methods. Table 4 presents the results of ablation experiments on the anomaly generator. The "Source Image" denotes the origin of the source image, where "Texture" represents the external texture dataset (DTD [3]) used by DRAEM, and "Fundus" represents the normal retinal images used for training. The "Self-Mix" indicates whether the Self-Mix module is applied to combine source and target information, and the "Mask P " indicates whether the Perlin mask P is used for anomaly generation.

The experimental results show that using an external texture dataset (DTD [3]) instead of consistent normal fundus images leads to a decline in performance. Additionally, the absence of the Self-Mix module results in a noticeable performance decrease, and excluding the Perlin mask P leads to a significant drop in performance. These findings underscore the importance of our proposed approach, which leverages consistent source images, diverse masks, and an appropriate fusion method can greatly improve performance.

4.4 Qualitative Results

As Shown in Fig. 4. It can be seen that the existing reconstruction-based method WDMT and synthetic anomalies based method DRAEM would detect the normal structure of the fundus (like vessel, optic cup and disc) to anomalies, as indicated by the red circles in the figure. Besides, the representation-based method ReSAD can not provide precise localization results, leading to more false positive (yellow) regions around the lesions. Compared to our method, the three baseline methods show more missing (red) areas and more false positive (yellow) areas around the normal structure or lesion in the fundus images. Moreover, all baseline methods will miss some subtle lesions which is hard to detect, while our method can detect various shapes and types of lesions better and provide relatively fine-grained localization results. Additionally, our method exhibits strong generalization capability on more types of lesions, as shown in the Supplementary.

5 Conclusion

In this paper, we introduced the ReSynthDetect network, a novel approach for unsupervised anomaly detection in fundus images. Our method incorporated a novel anomaly generator that produces consistent synthetic anomalies. Besides, we introduced a reconstruction network to extract reconstructive features, which were then fused with the localization network for synthetic anomaly localization. Our approach outperformed baseline methods on the IDRiD and EyeQ datasets, demonstrating its effectiveness in retina anomaly detection tasks.

References

- [1] Annisa Nurul Azhar and Masayu Leylia Khodra. Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis. *2020 7th International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2020*, pages 2980–2988, 2020. doi: 10.1109/ICAICTA49861.2020.9428882.
- [2] Christoph Baur, Stefan Denner, Benedikt Wiestler, Nassir Navab, and Shadi Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain mr images: a comparative study. *Medical Image Analysis*, 69:101952, 2021.
- [3] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3606–3613, 2014.
- [4] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021.
- [5] Will Cukierski Emma Dugas, Jared Jorge. Diabetic retinopathy detection, 2015. URL <https://kaggle.com/competitions/diabetic-retinopathy-detection>.
- [6] Pedro F Felzenszwalb and Daniel P Huttenlocher. Distance transforms of sampled functions. *Theory of computing*, 8(1):415–428, 2012.

- [7] Huazhu Fu, Boyang Wang, Jianbing Shen, Shanshan Cui, Yanwu Xu, Jiang Liu, and Ling Shao. Evaluation of retinal image quality assessment networks in different color-spaces. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 48–56. Springer, 2019.
- [8] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1705–1714, 2019.
- [9] Yijin Huang, Weikai Huang, Wenhao Luo, and Xiaoying Tang. Lesion2void: unsupervised anomaly detection in fundus images. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.
- [10] Ziyang Huang, Haoyu Wang, Jin Ye, Jingqi Niu, Can Tu, Yuncheng Yang, Shiyi Du, Zhongying Deng, Lixu Gu, and Junjun He. Revisiting nnu-net for iterative pseudo labeling and efficient sliding window inference. In *Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation: MICCAI 2022 Challenge, FLARE 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings*, pages 178–189. Springer, 2023.
- [11] Hendrik Blockeel Kristian Kersting and Siegfried Nijssen Filip Železný. Machine learning and knowledge discovery in databases. 2013.
- [12] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021.
- [13] Li Lin, Meng Li, Yijin Huang, Pujin Cheng, Honghui Xia, Kai Wang, Jin Yuan, and Xiaoying Tang. The sustech-sysu dataset for automated exudate detection and diabetic retinopathy grading. *Scientific Data*, 7(1):409, 2020.
- [14] Yifan Mao, Fei-Fei Xue, Ruixuan Wang, Jianguo Zhang, Wei-Shi Zheng, and Hongmei Liu. Abnormality detection in chest x-ray images using uncertainty prediction autoencoders. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 529–538. Springer, 2020.
- [15] Jingqi Niu, Shiwen Dong, Qinji Yu, Kang Dang, and Xiaowei Ding. Region and Spatial Aware Anomaly Detection for Fundus Images. pages 2–6, 2023. URL <http://arxiv.org/abs/2303.03817>.
- [16] Khalil Ouardini, Huijuan Yang, Balagopal Unnikrishnan, Manon Romain, Camille Garcin, Houssam Zenati, J Peter Campbell, Michael F Chiang, Jayashree Kalpathy-Cramer, Vijay Chandrasekhar, et al. Towards practical unsupervised anomaly detection on retinal images. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 225–234. Springer, 2019.
- [17] Ken Perlin. An image synthesizer. *ACM Siggraph Computer Graphics*, 19(3):287–296, 1985.

- [18] Prasanna Porwal, Samiksha Pachade, Manesh Kokare, Girish Deshmukh, Jaemin Son, Woong Bae, Lihong Liu, Jianzong Wang, Xinhui Liu, Liangxin Gao, et al. Idris: Diabetic retinopathy–segmentation and grading challenge. *Medical image analysis*, 59:101561, 2020.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [20] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022.
- [21] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.
- [22] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14902–14912, 2021.
- [23] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019.
- [24] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 474–489. Springer, 2022.
- [25] Satya P Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan, and Balázs Gulyás. 3d deep learning on medical images: a review. *Sensors*, 20(18): 5097, 2020.
- [26] Vitjan Zavrtanik, Matej Kristan, and Danijel Škočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021.
- [27] Wentian Zhang, Xu Sun, Yuexiang Li, Haozhe Liu, Nanjun He, Feng Liu, and Yefeng Zheng. A multi-task network with weight decay skip connection training for anomaly detection in retinal fundus images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 656–666. Springer, 2022.
- [28] He Zhao, Yuexiang Li, Nanjun He, Kai Ma, Leyuan Fang, Huiqi Li, and Yefeng Zheng. Anomaly detection for medical images using self-supervised and translation-consistent features. *IEEE Transactions on Medical Imaging*, 40(12):3641–3651, 2021.

- [29] Kang Zhou, Yuting Xiao, Jianlong Yang, Jun Cheng, Wen Liu, Weixin Luo, Zaiwang Gu, Jiang Liu, and Shenghua Gao. Encoding structure-texture relation with p-net for anomaly detection in retinal images. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pages 360–377. Springer, 2020.
- [30] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820–838, 2021.
- [31] Qikui Zhu, Yanqing Wang, Lei Yin, Jiancheng Yang, Fei Liao, and Shuo Li. Self-mix: A self-adaptive data augmentation method for lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, pages 683–692. Springer, 2022.
- [32] Karel Zuiderveld. Contrast limited adaptive histogram equalization. *Graphics gems*, pages 474–485, 1994.