A Multi-task Network with Weight Decay Skip Connection Training for Anomaly Detection in Retinal Fundus Images

Wentian Zhang[† 1,2], Xu Sun[† ✉ 2], Yuexiang Li[† 2], Haozhe Liu[2], Nanjun He[2], Feng Liu[✉ 1], and Yefeng Zheng[2]

[1] Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
`feng.liu@szu.edu.cn`
[2] Jarvis Lab, Tencent, Shenzhen, China
`ericxsun@tencent.com`

Abstract. By introducing the skip connection to bridge the semantic gap between encoder and decoder, U-shape architecture has been proven to be effective for recovering fine-grained details in dense prediction tasks. However, such a mechanism cannot be directly applied to reconstruction-based anomaly detection, since the skip connection might lead the model overfitting to an identity mapping between the input and output. In this paper, we propose a weight decay training strategy to progressively mute the skip connections of U-Net, which effectively adapts U-shape network to anomaly detection task. Thus, we are able to leverage the modeling capabilities of U-Net architecture, and meanwhile prevent the trained model from bypassing low-level features. Furthermore, we formulate an auxiliary task, namely histograms of oriented gradients (HOG) prediction, to encourage the framework to deeply exploit contextual information from fundus images. The HOG feature descriptors with three different resolutions are adopted as the auxiliary supervision signals. The multi-task framework is dedicated to enforce the model to aggregate shared significant commonalities and eventually improve the performance of anomaly detection. Experimental results on Indian Diabetic Retinopathy image Dataset (IDRiD) and Automatic Detection challenge on Age-related Macular degeneration dataset (ADAM) validate the superiority of our method for detecting abnormalities in retinal fundus images. The source code is available at https://github.com/WentianZhang-ML/WDMT-Net.

Keywords: Skip connection · Anomaly detection · Feature prediction · Fundus image.

† Equal Contribution
This work is done when Wentian Zhang is an intern at Jarvis Lab, Tencent.

1 Introduction

With the rapid development of artificial intelligence techniques in the past decades, deep supervised learning has proven its potential for automatic ocular disease screening or diagnosis using retinal fundus images [10,18]. However, training a highly accurate supervised classifier usually requires a fairly large amount of labeled data, which is extremely expensive and difficult to acquire due to the privacy issue of medical data. Even if the labeled data is available, the model often easily suffers from the class imbalance problem, as the data from healthy subjects is prevalent therefore easier to collect in a large quantity. For those reasons, anomaly detection, aiming to identify abnormalities with only normal images at the training stage, has drawn increasing attentions from the community [21,23,24]. Current anomaly detection methods can mainly be divided into two categories: the reconstruction-based and non-reconstruction-based methods. The former methods are established upon the assumption: the well-trained model can excellently reconstruct normal images while yield large reconstruction error for abnormal images. The latter ones rely on other techniques, such as transfer learning [12] and discriminative learning [5]. Compared to the non-reconstruction-based methods, the reconstruction-based ones are verified to achieve more robust anomaly detection performance. Hence, in this paper, we focus on the reconstruction-based anomaly detection for retinal fundus images.

By introducing skip connection in the encoder-decoder architecture, U-Net and its variants have achieved wide successes in biomedical image segmentation and image-to-image translation [7,8,9,14]. However, it is surprising to find that most existing reconstruction-based anomaly detection methods are built upon auto-encoder architecture without skip connections. Therefore, it remains an interesting question: *whether the skip connection can be helpful for improving the anomaly detection performance?* From another aspect, a recent research has revealed the effectiveness of histograms of oriented gradients (HOG) prediction for self-supervised representation learning [19]. For normal fundus images, large HOG values can be obtained from the areas around the blood vessels and optic disc, which contain the anatomical structure of the retina. Therefore, we raise a second question: *whether the HOG prediction task can serve the image reconstruction (main task) as auxiliary and assist the anomaly detection?*

To address the aforementioned two questions, we propose to train a multi-task encoder-decoder network with weight decay skip connection (WDMT-Net) for anomaly detection with retinal fundus images. The main contributions of this work can be summarized as follows. First, we explore the applicability of skip connection to reconstruction-based anomaly detection. Specifically, a weight decay skip connection training strategy is presented to mitigate the identity mapping problem of the U-Net architecture and meanwhile leverage its advantage on feature representation learning. Second, we integrate an auxiliary task, *i.e.,* HOG prediction, to the anomaly detection framework, which can fully exploit the significant commonalities of normal fundus images. Last but not least, our WDMT-Net outperforms the state-of-the-art methods on Indian Diabetic

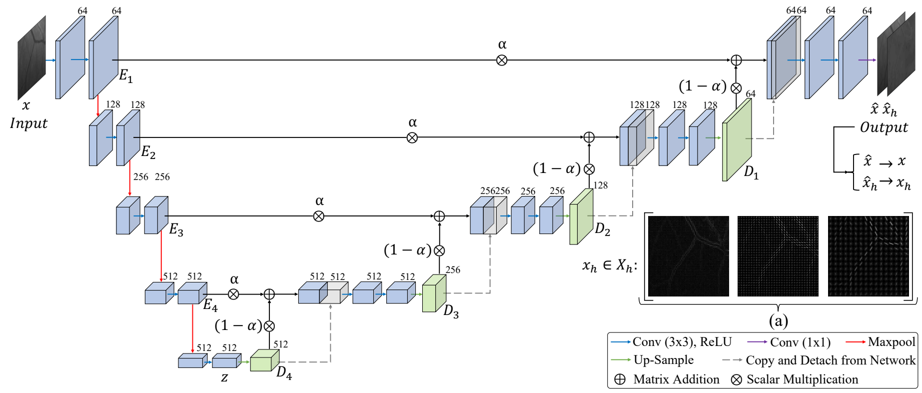

Fig. 1. The overall architecture of our proposed WDMT-Net for anomaly detection in retinal fundus images. (a) Examples of HOG features with different cell sizes: images from left to right are the HOG features obtained by cell sizes of 4×4, 8×8, and 16×16 pixels, respectively.

Retinopathy Image Dataset (IDRiD) [13], which demonstrates its effectiveness for detecting abnormal regions in retinal fundus images.

2 Method

Fig. 1 shows the overall architecture of our proposed WDMT-Net. The main components consist of a weight decay skip connection training strategy to leverage the modeling capabilities of U-Net architecture and a dual-output decoder to exploit shared commonalities between two related tasks (*i.e.*, image reconstruction and HOG prediction).

2.1 Weight Decay Skip Connection Training

By introducing skip connections to bridge the semantic gap between encoder and decoder, U-Net architecture has been proven to be effective in recovering fine-grained details of the target subjects. However, such a mechanism is rarely utilized in current reconstruction-based anomaly detection methods. The underlying reason may be that the skip connections at early stages tend to mislead the model to bypass the lower levels of features and essentially learn an identity mapping function. Such a dilemma significantly degrades the performance of U-Net for anomaly detection.

To mitigate this problem, we develop a simple-yet-effective weight decay training strategy to gradually mute the skip connections. Different from the original U-Net architecture, where the different levels of features in the encoder are directly concatenated to the corresponding decoded features, in WDMT-Net, we first formulate the weighted feature map at each spatial resolution as follows:

$$M_i = (\alpha \otimes E_i) \oplus ((1-\alpha) \otimes D_i), \tag{1}$$

where $\alpha \in [0, 1]$ is a weight factor; E_i and D_i represent the feature maps from the ith level of the encoder and decoder, respectively; $\oplus$ denotes feature addition; and $\otimes$ denotes scalar multiplication. Then, M_i is concatenated to a detached counterpart of D_i, which yields

$$D_i^{'} = Concat(M_i, \bar{D}_i), \tag{2}$$

where $Concat(\cdot)$ denotes the concatenation operation and $\bar{\cdot}$ denotes the detach operation. Note that we detach D_i from the network learning to restrict the gradients to propagate backward through the weighted feature path illustrated in Eq. (1).

During the training phase, the weight factor α in Eq. (1) is first initialized to 1, and then gradually decayed to 0. As shown in Fig. 1, when $\alpha = 1$, E_i is directly skip connected to $\bar{D}_i$ and our WDMT-Net is of the same structure to U-Net. However, since $\bar{D}_1$ is detached from the network learning flow, the computed gradients are not able to propagate from the upper layers to the lower layers in the decoder, which means only the network parameters at the first learning stage can be updated. As α decreases, the focus of optimization gradually varies from the horizontal decoder-encoder skip connection direction to the up-down direction. Therefore, the lower-level features learned at the early stage gradually aggregate to the higher-level features. Finally, when $\alpha = 0$, M_i is in fact a copy of D_i and our WDMT-Net degrades to an encoder-decoder network without skip connection. In the final network, D_i' contains two copies of D_i and this redundancy can be removed by re-organizing the weights.

2.2 HOG Prediction as an Auxiliary Task

As revealed by a recent research, HOG prediction could be an exceedingly effective way for self-supervised representation learning [19]. Meanwhile, large HOG values can be yielded around the blood vessels and optic disc, which provide the useful anatomical structure information for the representation learning of normal retinal fundus images. Based on such observation, we propose to formulate a multi-task network [15] to simultaneously regress the pixel intensity and the HOG feature of the input images, which enforces the model to learn the shared commonalities beneficial for anomaly detection. Let x denote an input image, the proposed WDMT-Net can then be formulated as:

$$\langle \hat{x}, \hat{x}_h \rangle = Dec(Enc(x)), \tag{3}$$

where the $Enc(\cdot)$ and $Dec(\cdot)$ are the encoder and decoder of the WDMT-Net, respectively; $\hat{x}$ is the reconstruction of the input image; and $\hat{x}_h$ is the predicted HOG feature map. The overall learning objective is defined as:

$$\mathcal{L} = \|\hat{x} - x\|_2^2 + \|\hat{x}_h - x_h\|_2^2, \tag{4}$$

where the training sample x is merely drawn from the normal images; the ground-truth of the corresponding HOG feature map $x_h \in X_h$ is randomly drawn from a pool X_h that contains HOG features computed with different cell resolutions; and $||\cdot||_2$ denotes the L_2-norm.

In essence, HOG is a feature descriptor, which describes the distribution of gradient orientations or edge directions over the local cells [3], and with different local cell size one can obtain HOG features in different scales. Fig. 1 (a) shows the computed HOG features with three different spatial cell sizes, which are used in the feature prediction label pool during the training phase. There are two reasons for this setting: First, the use of multi-scale HOG features reflects the fact that the internal structure of fundus images is of varying sizes. Second, the random selection of HOG target label works as an effective data augmentation strategy that increases the diversity of input-output data pairs being fed to the model.

2.3 Anomaly Detection

Similar to existing reconstruction-based anomaly detection methods, our WDMT-Net is built based on the assumption that the abnormal images cannot be well reconstructed by a model trained merely with the normal images. Taking a gray image x_t as an input at the test stage, the proposed method can reconstruct a new image $\hat{x}_t$. We compute the the anomaly score map for the pixel-level anomaly detection in the image space as

$$\mathcal{A}_M = \left|\hat{x}_t - x_t\right|. \tag{5}$$

The larger the reconstruction error is, the higher possibility of the corresponding region to be abnormal.

3 Experiments

3.1 Dataset and Implementation Details

In this section, we adopt two publicly available datasets, namely Indian Diabetic Retinopathy Image Dataset (IDRiD) [13] and Automatic Detection challenge on Age-related Macular degeneration dataset (ADAM) [4], for performance evaluation. Following [23], we only choose the normal class from the original training set to train the proposed model and use the lesion detection dataset as the test set. In IDRiD dataset, there are 134 normal images for training, and 81 abnormal images for testing. The pixel-level annotation of abnormal images contains four different lesions, including haemorrhages, microaneurysms, hard and soft exudates. For ADAM dataset, it contains 282 normal images and 118 abnormal images with five different lesions and corresponding pixel-level annotations, including drusen, exudate, hemorrhage, scars and other lesions. Since the original images of both dataset are very large (*i.e.*, 4,288×2,848 pixels and 2124×2056 pixels), we resize each image to 768×768 pixels and then crop them into 3×3

Table 1. Ablation study of our WDMT-Net. SC, WD, and HOG represent the use of skip connection, weight decay training strategy and HOG prediction, respectively.

Model	Combination			IDRiD [13]			ADAM [4]		
	SC	WD	HOG	AUC	ACC	F1-score	AUC	ACC	F1-score
Auto-Encoder [2]				0.686	0.627	0.537	0.659	0.637	0.469
U-Net [14]	✓			0.553	0.564	0.532	0.610	0.619	**0.530**
WDMT-Net w/o HOG	✓	✓		0.725	0.667	0.680	0.670	0.654	0.484
Auto-Encoder			✓	0.715	0.655	0.664	0.662	0.643	0.470
U-Net	✓		✓	0.640	0.597	0.539	0.656	0.641	0.451
WDMT-Net (*Ours*)	✓	✓	✓	**0.748**	**0.694**	**0.711**	**0.687**	**0.660**	0.474

non-overlapping patches. After that, we transform each patch to gray-scale image for training and test. To supervise the HOG prediction task, we extract the HOG features with three different cell sizes, as mentioned in Section 2.2.

In our implementation, the proposed method is trained by the Adam optimizer with a learning rate of 1×10^{-4} and a weight decay of 5×10^{-5}. All code is implemented with PyTorch on a single NVIDIA RTX 3090 GPU with 24GB of memory and the batch size is set as 32. For simplicity, the weight factor α is initialized to 1 and decayed by Δ per epoch, where Δ is set as 0.05 by default. The decaying procedure stops when α reaches 0. Following previous work [23], the anomaly detection results are evaluated quantitatively by the the area under the curve (AUC), balanced accuracy (ACC), and F1-score.

3.2 Ablation Study

We perform an ablation study to investigate the contribution made by different components of the proposed WDMT-Net to anomaly detection in retinal fundus images. To make a fair comparison, the network architectures of Auto-Encoder [2] and U-Net [14] are set the same as our WDMT-Net, except for the setting of skip connection. It is worthwhile to mention that models under different settings are trained with the same protocol stated in Section 3.1,

Weight Decay Skip Connection Training. As shown in Table 1, the performance of U-Net is worse than Auto-Encoder, no matter with or without the auxiliary HOG prediction task. This indicates that directly applying skip connection to encoder-decoder network deteriorates the performance of anomaly detection. Nevertheless, by using the proposed weight decay skip connection training strategy, WDMT-Net consistently achieves better results with and without HOG prediction, which indicates the effectiveness of the proposed training strategy.

Moreover, in order to illustrate the role of skip connection in anomaly detection network, we visualize the image reconstruction loss versus training epoch on IDRiD dataset in Fig. 2. From the reconstruction loss curves, we observe that U-Net consistently achieves the lowest loss, which means skip connections do enable U-Net to reconstruct the training images more accurately. However, the same rule does not hold for anomaly detection at the test stage. Thus, we can conclude that the naive utilization of skip connections do raise the issue of

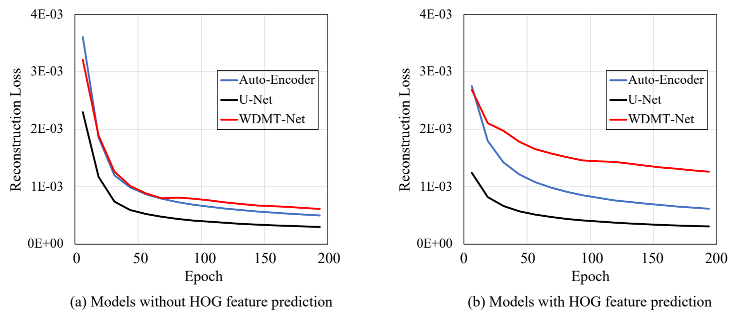

Fig. 2. The image reconstruction loss vs. training epochs on IDRiD dataset.

Table 2. Impact of the decay rate Δ of the skip connection in our WDMT-Net.

Decay setting	IDRiD [13]			ADAM [4]		
	AUC	ACC	F1-score	AUC	ACC	F1-score
$\Delta = 0.005$	0.729	0.661	0.687	0.676	0.663	0.451
$\Delta = 0.01$	0.738	0.685	0.692	0.678	0.662	**0.482**
$\Delta = 0.025$	0.731	0.674	0.680	0.673	**0.663**	0.465
$\Delta = 0.05$	**0.748**	**0.694**	**0.711**	**0.687**	0.660	0.474
$\Delta = 0.1$	0.724	0.669	0.709	0.674	0.660	0.471

identity mapping. In contrast, our WDMT-Net can mitigate this problem by using the weight decay training strategy.

The setting of hyper-parameters is also an important factor, which may affect the model performance. To this end, we conduct experiments to evaluate the model performance with different decay rate of the skip connections. As shown in Table 2, we set Δ as 0.005, 0.01, 0.025, 0.05 and 0.1 for comparison. It can be observed that our WDMT-Net model with $\Delta = 0.05$ achieves the best results.

Multi-task Learning. To extensively evaluate the contribution of HOG prediction for anomaly detection, we also apply the same multi-task learning scheme for Auto-Encoder and U-Net. Due to the extra information provided by the HOG prediction task, the anomaly detection performances of all multi-task models are consistently improved as shown in Table 1.

The proposed method learns the multi-resolution HOG features from a label pool that contains HOG features obtained by cell sizes of 4×4, 8×8, and 16×16 pixels, respectively. To verify the effectiveness of such a setting, we perform the following experiments based on WDMT-Net. First, we formulate three models where each of them only uses a single-scale of HOG features as the learning target. As shown in Table 3, among these three models, the one learned with HOG of the 16×16 cell size obtains the best results, However, it still underperforms our proposed method, which indicates the advantage of using multi-scale targets in WDMT-Net. Moreover, we also compare the performance of WDMT-Net with two alternative models in which the multi-scale HOG features are utilized in dif-

Table 3. Impact of auxiliary HOG prediction task setting of the proposed WDMT-Net.

Setting of the prediction target	IDRiD [13]			ADAM [4]		
	AUC	ACC	F1-score	AUC	ACC	F1-score
HOG features with 4×4 cells	0.736	0.668	0.704	0.677	0.653	0.468
HOG features with 8×8 cells	0.736	0.674	0.696	0.682	0.649	0.460
HOG features with 16×16 cells	0.738	0.671	0.685	0.682	0.660	**0.493**
Three HOG features as three outputs	0.731	0.678	**0.720**	0.660	0.647	0.439
The average of three HOG features	0.733	0.669	0.700	0.648	0.631	0.395
Three HOG features as a label pool	**0.748**	**0.694**	0.711	**0.687**	**0.660**	0.474

Table 4. Quantitative comparison of the proposed WDMT-Net with the state-of-the-art methods.

Method	IDRiD [13]			ADAM [4]		
	AUC	ACC	F1-score	AUC	ACC	F1-score
Auto-Encoder [2]	0.686	0.627	0.537	0.659	0.637	0.469
MemAE [6]	0.647	0.592	0.567	0.667	0.647	0.439
BiO-Net [20]	0.606	0.563	0.519	0.642	0.612	0.481
Attn U-Net [11]	0.581	0.555	0.558	0.645	0.617	0.408
AnoGAN [17]	0.630	0.618	0.579	0.677	**0.661**	0.455
f-AnoGAN [16]	0.698	0.686	0.637	0.662	0.638	0.455
GANomaly [1]	0.652	0.633	0.658	0.673	0.618	**0.539**
Sparse-GAN [22]	0.663	0.638	0.651	0.667	0.627	0.500
ProxyAno [23]	0.701	0.682	0.649	0.675	0.648	0.451
WDMT-Net (*Ours*)	**0.748**	**0.694**	**0.711**	**0.687**	0.660	0.474

ferent ways (*i.e.*, to predict HOG at different resolutions simultaneously, and to predict the average of them). As shown in Table 3, our method outperforms these two alternative settings too in terms of AUC and ACC, which further indicates the effectiveness of our random sampling method.

3.3 Comparison to State-of-the-art Methods

To further validate the superiority of our method, we compare our method with several state-of-the-art anomaly detection methods, including Auto-Encoder [2], MemAE [6], Attn U-Net [11], BiO-Net [20], AnoGAN [17], f-AnoGAN [16], GANomaly [1], Sparse-GAN [22], and ProxyAno [23]. It can be seen that, the proposed WDMT-Net outperforms the state-of-the-art methods. We further provide some results in Fig. 3. It can be seen that, the normal patches can be reconstructed with a small error, while the abnormal patches (*i.e.*, patches with lesion) are reconstructed with a large error. The prediction $\mathcal{A}_M$ matches the pixel-level lesion ground truth pretty well. These results further validate the superiority of our method.

4 Conclusion

In the work, we explored the applicability of skip connection and multi-task learning to anomaly detection tasks. Concretely, a weight decay training strategy was proposed to effectively adapt U-shape network for the anomaly detection

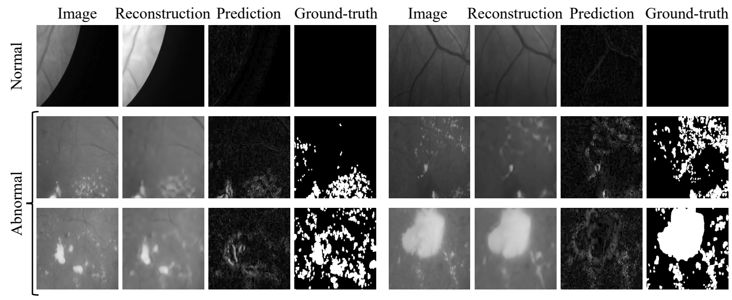

Fig. 3. The qualitative results of WDMT-Net on IDRiD retinal fundus images.

task, which prevented the model from overfitting to the identity mapping introduced by skip connections. Furthermore, an auxiliary task, *i.e.,* HOG prediction, was integrated to our framework to explore the effectiveness of multi-task learning. Such a multi-task framework was dedicated to enforce the model to aggregate shared commonalities between these two tasks and finally improve the performance of anomaly detection. Extensive experiments on publicly available IDRiD and ADAM fundus image datasets demonstrated the superiority of our framework to the state-of-the-art anomaly detection methods. In the future, we plan to expand the applicability of our WDMT-Net to more medical imaging modalities.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (Grant 62076163 and Grant 91959108), the Shenzhen Fundamental Research Fund (Grant JCYJ20190808163401646), Key-Area Research and Development Program of Guangdong Province, China (No. 2018B010111001), National Key R&D Program of China (2018YFC2000702) and the Scientific and Technical Innovation 2030-"New Generation Artificial Intelligence" Project (No. 2020AAA0104100).

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: Semi-supervised anomaly detection via adversarial training. In: Asian Conference on Computer Vision. pp. 622–637. Springer (2018)
2. Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In: International MICCAI Brainlesion Workshop. pp. 161–169. Springer (2018)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 886–893. Ieee (2005)
4. Fang, H., Li, F., Fu, H., Sun, X., Cao, X., Lin, F., Son, J., Kim, S., Quellec, G., Matta, S., et al.: ADAM challenge: Detecting age-related macular degeneration from fundus images. IEEE Transactions on Medical Imaging (2022)

5. Golan, I., El-Yaniv, R.: Deep anomaly detection using geometric transformations. Advances in Neural Information Processing Systems **31** (2018)
6. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1705–1714 (2019)
7. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: Context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging **38**(10), 2281–2292 (2019)
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)
9. Ji, W., Yu, S., Wu, J., Ma, K., Bian, C., Bi, Q., Li, J., Liu, H., Cheng, L., Zheng, Y.: Learning calibrated medical image segmentation via multi-rater agreement modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12341–12351 (2021)
10. Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., Fu, H.: Applications of deep learning in fundus images: A review. Medical Image Analysis **69**, 101971 (2021)
11. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
12. Ouardini, K., Yang, H., Unnikrishnan, B., Romain, M., Garcin, C., Zenati, H., Campbell, J.P., Chiang, M.F., Kalpathy-Cramer, J., Chandrasekhar, V., 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, pp. 225–234. Springer (2019)
13. Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., et al.: IDRiD: Diabetic retinopathy–segmentation and grading challenge. Medical Image Analysis **59**, 101561 (2020)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 234–241. Springer (2015)
15. Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 (2017)
16. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis **54**, 30–44 (2019)
17. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. pp. 146–157. Springer (2017)
18. Son, J., Shin, J.Y., Kim, H.D., Jung, K.H., Park, K.H., Park, S.J.: Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. Ophthalmology **127**(1), 85–94 (2020)
19. Wei, C., Fan, H., Xie, S., Wu, C.Y., Yuille, A., Feichtenhofer, C.: Masked feature prediction for self-supervised visual pre-training. arXiv preprint arXiv:2112.09133 (2021)
20. Xiang, T., Zhang, C., Liu, D., Song, Y., Huang, H., Cai, W.: BiO-Net: learning recurrent bi-directional connections for encoder-decoder architecture. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 74–84. Springer (2020)

21. Zhao, H., Li, Y., He, N., Ma, K., Fang, L., Li, H., Zheng, Y.: Anomaly detection for medical images using self-supervised and translation-consistent features. IEEE Transactions on Medical Imaging **40**(12), 3641–3651 (2021)
22. Zhou, K., Gao, S., Cheng, J., Gu, Z., Fu, H., Tu, Z., Yang, J., Zhao, Y., Liu, J.: Sparse-GAN: Sparsity-constrained generative adversarial network for anomaly detection in retinal OCT image. In: IEEE 17th International Symposium on Biomedical Imaging. pp. 1227–1231. IEEE (2020)
23. Zhou, K., Li, J., Luo, W., Li, Z., Yang, J., Fu, H., Cheng, J., Liu, J., Gao, S.: Proxy-bridged image reconstruction network for anomaly detection in medical images. IEEE Transactions on Medical Imaging (2021)
24. Zimmerer, D., Isensee, F., Petersen, J., Kohl, S., Maier-Hein, K.: Unsupervised anomaly localization using variational auto-encoders. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 289–297. Springer (2019)