A Model-Agnostic Framework for Universal Anomaly Detection of Multi-organ and Multi-modal Images

Yinghao Zhang[1], Donghuan Lu[2], Munan Ning[3], Liansheng Wang[1(✉)], Dong Wei[2(✉)], and Yefeng Zheng[2]

[1] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China
lswang@xmu.edu.cn
[2] Tencent Healthcare Co., Jarvis Lab, Shenzhen, China
donwei@tencent.com
[3] Peking University, Shenzhen Graduate School, Shenzhen, China

Abstract. The recent success of deep learning relies heavily on the large amount of labeled data. However, acquiring manually annotated symptomatic medical images is notoriously time-consuming and laborious, especially for rare or new diseases. In contrast, normal images from symptom-free healthy subjects without the need of manual annotation are much easier to acquire. In this regard, deep learning based anomaly detection approaches using only normal images are actively studied, achieving significantly better performance than conventional methods. Nevertheless, the previous works committed to develop a specific network for each organ and modality separately, ignoring the intrinsic similarity among images within medical field. In this paper, we propose a model-agnostic framework to detect the abnormalities of various organs and modalities with a single network. By imposing organ and modality classification constraints along with center constraint on the disentangled latent representation, the proposed framework not only improves the generalization ability of the network towards the simultaneous detection of anomalous images with various organs and modalities, but also boosts the performance on each single organ and modality. Extensive experiments with four different baseline models on three public datasets demonstrate the superiority of the proposed framework as well as the effectiveness of each component.

Keywords: Anomaly Detection · Medical Images · Multi-Organ · Multi-modality

1 Introduction

Although deep learning has achieved great success in various computer vision tasks [11], the requirement of large amounts of labeled data limits its application in

Y. Zhang and D. Lu—Contributed equally.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43898-1_23.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 232–241, 2023.
https://doi.org/10.1007/978-3-031-43898-1_23

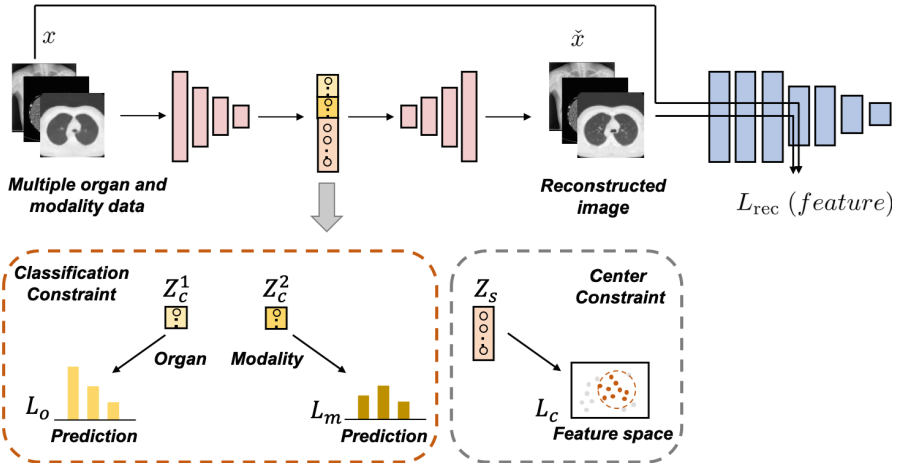

Fig. 1. Overview of the proposed framework incorporated into the DPA method [15]. Two classification constraints (organ and modality), and a center constraint are applied on the disentangled latent representation in addition to the original loss(es) of the baseline model.

the field of medical image analysis. Annotated abnormal images are difficult to acquire, especially for rare or new diseases, such as the COVID-19, whereas normal images are much easier to obtain. Therefore, many efforts [13–16] have been made on deep learning based medical anomaly detection, which aims to learn the distribution of normal patterns from healthy subjects and detect the anomalous ones as outliers.

Due to the absence of anomalous subjects, most previous studies adopted the encoder-decoder structure or generative adversarial network (GAN) as backbone to obtain image reconstruction error as the metric for recognition of outliers [5,13,20]. Other approaches [14–16] learned the normal distribution[1] along with the decision boundaries to differentiate anomalous subjects from normal ones. Despite the improvement they achieved over traditional one-class classification methods [1,9,12], all these works committed to train a specific network to detect the anomalies of each organ, neglecting the intrinsic similarity among different organs. We hypothesize that there are underlying patterns in the normal images of various organs and modalities despite their seemly different appearance, and a model, which can fully exploit their latent information, not only has better generalization ability to recognize the anomalies within different organs/modalities, but also can achieve superior performance for detecting the anomalies of each of them.

To this end, we propose a novel model-agnostic framework, denoted as Multi-Anomaly Detection with Disentangled Representation (MADDR), for the simultaneous detection of anomalous images within different organs and modalities. As displayed in Fig. 1, to fully explore the underlying patterns of normal images as well as bridge the appearance gap among different organs and modalities, the latent representation z is

[1] Note in this work we use the term 'normal distribution' to refer to the distribution of images of normal, healthy subjects, instead of the Gaussian distribution.

disentangled into three parts, i.e., two categorical parts (z_o and z_m, corresponding to organ and modality, respectively) and a continuous variable (z_c). The first two parts represent the categorical information for the distinction of specific organs and modalities, respectively, while the last part denotes the feature representation for characterizing each individual subject. Specifically, we propose to impose an organ classification constraint (L_o) as well as a modality classification constraint (L_m) on the categorical parts to leverage the categorical information, and a center constraint on the continuous variable part to compact the feature representation of the normal distribution so that the outliers can be easier to identify. It is worth mentioning that the categorical label of medical images are easy to obtain because such information should be recorded during image acquisition, and with the disentanglement strategy, the potential contradiction between the classification constraint (which aims to separate images of different organs and different modalities) and the center constraint (which tends to compact the feature representation) can be avoided. Our contributions can be summarized as follows:

- To the best of our knowledge, this is the first study to detect the anomalies of multiple organs and modalities with a single network. We show that introducing images from different organs and modalities with the proposed framework not only extends the generalization ability of the network towards the recognition of the anomalies within various data, but also improves its performance on each single kind.
- We propose to disentangle the latent representation into three parts so that the categorical information can be fully exploited through the classification constraints, and the feature representation of normal images is tightly clustered with the center constraint for better identification of anomalous pattern.
- Extensive experiments demonstrate the superiority of the proposed framework regarding the medical anomaly detection task, as well as its universal applicability to various baseline models. Moreover, the effectiveness of each component is evaluated and discussed with thorough ablation study.

2 Methodology

Unlike previous approaches which trained a separate model to capture the anomalies for each individual organ and modality, in this study we aim to exploit the normal images of multiple organs and modalities to train a generic network towards better anomaly detection performance for each of them. The proposed framework is model-agnostic and can be readily applied to most standard anomaly detection methods. For demonstration, we adopt four state-of-the-art anomaly detection methods, i.e., deep perceptual autoencoder (DPA) [15], memory-augmented autoencoder (MemAE) [4], generative adversarial networks based anomaly detection (GANomaly) [2], and fast unsupervised anomaly detection with generative adversarial networks (f-AnoGAN) [13] as baseline methods. In this section, we present the proposed universal framework for the anomaly detection task of medical images in details.

2.1 Framework Overview

We first formulate the anomaly detection task for images with various organs and modalities. For a dataset targeting a specific organ k ($k \in [0, K]$) and modality l

($l \in [0, L]$), there is a training dataset $D_{k,l}$ with only normal images and a test set $D^t_{k,l}$ with both normal and abnormal images. In this study, we use N_k to represent total number of images targeting organ k and N_l to represent total number of images belong to modality l. The goal of our study is to train a generic model with these multi-organ and multi-modality normal images to capture the intrinsic normal distribution of training sets, such that the anomalies in test sets can be recognized as outliers.

To better elaborate the process of applying our MADDR framework on standard anomaly detection methods, we first briefly introduce the workflow of a baseline method, DPA [15]. As shown in Fig. 1, the network of DPA consists of an autoencoder and a pre-trained feature extractor. Through autoencoder, the images are encoded into latent representations and then reconstructed into the original image space. The relative perceptual loss is adopted as the objective function for the optimization of autoencoder and the measurement of anomaly, which is defined as $L_{\text{rec}}(x, \tilde{x}) = \frac{\|\hat{f}(x) - \hat{f}(\tilde{x})\|_1}{\|\hat{f}(x)\|_1}$, where x and $\tilde{x}$ denote the input image and the reconstructed one, respectively. $\hat{f}(x) = \frac{f(x)-\mu}{\sigma}$ is the normalized feature with mean μ and standard deviation σ pre-calculated on a large dataset, where $f(\cdot)$ represents the mapping function of the pre-trained feature extractor. By comparing the features of original images and the reconstructed ones through the relative perceptual loss, the subjects with loss larger than the threshold are recognized as abnormal ones. To fully exploit the underlying patterns in the normal images of various organs and modalities, we incorporate additional constraints on the encoded latent representations.

Specifically, our MADDR approach encourages the model to convert the input image x into a latent representation z, which consists of disentangled category and individuality information. To be more precise, the encoded latent representation z is decomposed into three parts, i.e., the organ category part z_o, the modality category part z_m and the continuous variable part z_c. Here, z_o and z_m represent the categorical information (which is later converted into the probabilities of x belonging to each organ and modality through two separate fully-connected layers), and z_c denotes the feature representation for characterizing each individual image (which should be trained to follow the distribution of normal images). Leveraging the recorded categorical information of the images, we impose two classification constraints on z_o and z_m, respectively, along with a center constraint on z_c to compact the cluster of feature representation in addition to the original loss(es) of the baseline methods. In this study, we evaluate the proposed model-agnostic framework on four cutting-edge methods [2,4,13,15] with their networks as baseline models, and their original losses along with the proposed constraints as the training objective functions. If we use L_b, L_o, L_m and L_c to represent the loss of the baseline method, the organ classification constraint, the modality classification constraint and the center constraint, respectively, the overall loss function of the proposed framework can be formulated as:

$$L = L_b(X) + \lambda_1 L_o(X, Y) + \lambda_2 L_m(X, Y) + \lambda_3 L_c(X), \tag{1}$$

where X and Y denote the set of images and labels, respectively, while λ_1, λ_2 and λ_3 are the weights to balance different losses, and set to 1 in this study (results of exploratory experiment are displayed in the supplementary material).

2.2 Organ and Modality Classification Constraints

Benefiting from the acquisition procedure of medical images, the target organ and modality of each scan should be recorded and can be readily used in our study. Based on the assumption that a related task could provide auxiliary guidance for network training towards superior performance regrading the original task, we introduce two additional classification constraints (organ and modality classification) to fully exploit such categorical information. Through an additional organ classifier (a fully-connected layer), the organ classification constraint is applied on the transformed category representation z_o by distinguishing different organs. A similar constraint is also applied on the modality representation z_m. Considering the potential data imbalance issue among images of different organs and modalities, we adopt the focal loss [7] as the classification constraints by adding a modulating factor to the cross entropy loss.

Using z_o^i to represent the organ category representation of image i, the organ classification loss can be formulated as:

$$L_o = -\sum_{k=1}^{K}\sum_{i=1}^{N_k} \alpha_o^k \left(1 - P^k(z_o^i)\right)^\gamma \log\left(P^k(z_o^i)\right), \tag{2}$$

where α_o^k denotes the weight to balance the impact of different organs and $P^k(z_o^i)$ represents the probability of image i belonging to class k. The focusing parameter γ can reduce the contribution of easy samples to the loss function and extend the range of loss values for comparison.

Similarity, with the modality category representation z_m^i, the modality classification loss can be written as:

$$L_m = -\sum_{l=1}^{L}\sum_{i=1}^{N_l} \alpha_m^l \left(1 - P^l(z_m^i)\right)^\gamma \log\left(P^l(z_m^i)\right). \tag{3}$$

2.3 Center Constraint

Intuitively, the desired quality of a representation is to have similar feature embeddings for images of the same class. Because in the anomaly detection task, all the images used for training belong to the same normal group, we impose a center constraint so that the features from the normal images are tightly clustered to the center and the encoded features of abnormal images lying far from the normal cluster are easy to identify. However, directly compacting the latent representation into a cluster is potentially contradictory to the organ and modality classification tasks which aim to separate different organs and modalities. To avoid the contradiction, we propose to impose the center constraint only on the continuous variable part z_c with Euclidean distance as the measurement of the compactness. Similar to [10], if we use z_c^i to represent the continuous variable representation of image i, the measurement of compactness can be defined as:

$$L_c = \sum_{b=1}^{B} \frac{1}{L_b} \sum_{i=1}^{L_b} (z_c^i - m^i)^T (z_c^i - m^i), \tag{4}$$

where L_b represents the number of images in batch b, B denotes the number of batches and $m^i = \frac{1}{L_b - 1} \sum_{j \neq i} z_c^j$ is the mean of the rest images in the same batch as image i.

2.4 Optimization and Inference

To demonstrate the effectiveness of the proposed framework, we inherit the network architectures and most procedures of the baseline methods for a fair comparison. During the optimization stage, there are only two differences: 1) we introduce three additional losses as stated above; 2) considering the limited number of images, mixup [17] is applied for data augmentation to better bridge the gap among different organs and modalities. For other factors, such as optimization algorithm and related hyper-parameters, we follow the original settings of the baseline methods. During inference, the same metrics of the baseline methods are adopted to measure the anomaly scores. For the details about the baseline methods, please refer to their original studies [2,4,13,15].

3 Experiments

3.1 Experimental Setting

We evaluate our method on three benchmark datasets targeting various organs and modalities as stated below.

- The LiTS-CT dataset [3] consists of 3D volumetric data with rare healthy subjects. To ensure a sufficient amount of training data, we remove the 2D abnormal slices from some patients and use the rest normal slices for training. Therefore, 186 axial CT slices from 77 3D volumes are used as normal data for training, 192 normal and 105 abnormal slices from 27 3D volumes for validation, and 164 normal and 249 abnormal slices from the rest 27 3D volumes for testing.
- The Lung-X-rays [6] is a small dataset consists of 2D frontal chest X-ray images primarily from a hospital clinical routine. Following the same data split protocol provided by [19], the images are divided into three groups for training (228 normal images), validation (33 normal and 34 abnormal images) and testing (65 normal and 67 abnormal images).
- The Lung-CT [18] contains 2D slices regarding COVID-19. An official data split is provided, which contains 234 normal images for training, 58 normal and 60 abnormal images for validation, and 105 normal and 98 abnormal images for testing.

To evaluate the performance of proposed framework, we adopt three widely used metrics, including the area under the curve (AUC) of the receiver operating characteristic, F1-score and accuracy (ACC). Grid search is performed to find the optimal threshold based on the F1-score of the validation set. For all three metrics, a higher score implies better performance. The experiments are repeated three times with different random seeds to verify the robustness of the framework and provide more reliable results.

The framework is implemented with PyTorch 1.4 toolbox [8] using an NVIDIA Titan X GPU. As detailed in the supplementary material, we keep most parameters the same as the baseline methods, except for the original dimension of DPA's latent feature, which is increased from 16 to 128 to ensure sufficient model capacity for valid latent representation disentanglement of each image. For the hyper-parameters of focal loss, both α_o^k and α_m^l are set as 0.25, and γ is set to 2. For preprocessing, we first

Table 1. Anomaly detection performance on LiTS, Lung-X-rays and Lung-CT datasets. Mean and standard deviations of three metrics are presented.

Test-Datasets	Methods	ACC (%)	F1-score (%)	AUC (%)
LiTS [3]	DPA [15]	56.23 ± 2.67	55.92 ± 3.71	61.34 ± 4.64
	MADDR + DPA	64.87 ± 5.11	64.85 ± 5.36	72.38 ± 1.99
	MemAE [4]	51.40 ± 3.04	50.58 ± 4.14	55.96 ± 3.16
	MADDR + MemAE	51.63 ± 3.91	51.23 ± 3.58	57.73 ± 6.96
	f-AnoGAN [13]	59.26 ± 3.21	58.77 ± 1.89	55.90 ± 5.09
	MADDR + f-AnoGAN	59.93 ± 4.29	59.64 ± 2.66	61.77 ± 6.40
	GANomaly [2]	64.76 ± 6.06	63.93 ± 7.12	72.54 ± 2.09
	MADDR + GANomaly	75.64 ± 2.44	76.03 ± 2.43	79.20 ± 9.87
Lung-X-rays [6]	DPA [15]	64.14 ± 0.43	63.98 ± 0.35	72.23 ± 2.25
	MADDR + DPA	66.92 ± 1.16	66.30 ± 1.80	75.01 ± 0.95
	MemAE [4]	67.17 ± 1.16	66.40 ± 1.20	73.62 ± 3.09
	MADDR + MemAE	67.93 ± 3.82	67.39 ± 4.19	76.27 ± 3.53
	f-AnoGAN [13]	54.29 ± 3.81	52.78 ± 3.12	57.45 ± 5.48
	MADDR + f-AnoGAN	57.77 ± 7.95	56.22 ± 6.34	66.19 ± 7.59
	GANomaly [2]	57.71 ± 2.28	56.30 ± 3.20	64.97 ± 2.32
	MADDR + GANomaly	64.68 ± 6.89	63.13 ± 9.04	68.45 ± 4.67
Lung-CT [18]	DPA [15]	57.31 ± 1.42	57.09 ± 1.27	56.17 ± 0.87
	MADDR + DPA	59.28 ± 3.20	58.90 ± 3.12	58.86 ± 0.86
	MemAE [4]	51.40 ± 4.19	50.10 ± 5.86	52.67 ± 5.04
	MADDR + MemAE	55.67 ± 2.47	54.64 ± 3.36	57.13 ± 2.45
	f-AnoGAN [13]	56.65 ± 5.49	55.85 ± 6.54	58.13 ± 1.59
	MADDR + f-AnoGAN	58.17 ± 2.80	57.60 ± 3.14	61.30 ± 3.03
	GANomaly [2]	62.73 ± 1.99	61.61 ± 2.19	67.74 ± 1.15
	MADDR + GANomaly	67.16 ± 4.93	66.37 ± 5.99	69.91 ± 2.39

resize each image to 256×256 pixels, and then crop to 224×224 pixels. The data augmentation methods are applied on the fly during training, including random cropping, random rotation, mixup and random flipping. The code can be found in https://github.com/lianjizhe/MADDR_code.

3.2 Comparison Study

The quantitative results of the proposed framework and the state-of-the-art methods are displayed in Table 1. In four experiments with different baseline methods, we can observe significant improvement in all three metrics on various organs and modalities, demonstrating that the proposed framework can effectively boost the anomaly detection performance regardless the baseline approaches. In addition, the baseline methods need to train separate networks for different organs and modalities, while with the

Table 2. Quantitative ablation study of MADDR + DPA [15] with AUC (%) as evaluation metric.

		Mix Datasets			Loss				Test Datasets			
Index	Method	LiTS	Lung-CT	Lung-X-rays	L_O	L_M	L_C	f.d.	z	LiTS [3]	Lung-CT [18]	Lung-X-rays [6]
1	Single dataset	×	×	×	×	×	×	×	16	58.11 ± 5.82	53.63 ± 1.19	71.93 ± 1.01
2	Single dataset	×	×	×	×	×	×	×	128	61.34 ± 4.64	56.17 ± 0.87	72.23 ± 2.25
3	Three datasets	✓	✓	✓	×	×	×	×	128	59.75 ± 3.61	55.83 ± 1.37	69.14 ± 1.66
4	Multi-organ	✓	✓	×	✓	×	×	×	128	61.80 ± 4.27	56.19 ± 2.24	–
5	Multi-organ	✓	✓	×	✓	×	✓	×	128	61.81 ± 1.86	56.56 ± 2.37	–
6	Multi-organ	✓	✓	×	✓	×	✓	✓	128	62.49 ± 0.96	56.96 ± 1.61	–
7	Multi-modality	×	✓	✓	×	✓	×	×	128	–	56.65 ± 2.15	73.03 ± 1.33
8	Multi-modality	×	✓	✓	×	✓	✓	×	128	–	56.18 ± 0.79	73.45 ± 1.20
9	Multi-modality	×	✓	✓	×	✓	✓	✓	128	–	57.74 ± 1.75	74.28 ± 0.66
10	Multi-organ and -modality	✓	✓	✓	✓	✓	×	×	128	69.11 ± 7.23	57.38 ± 1.89	73.04 ± 0.40
11	Multi-organ and -modality	✓	✓	✓	✓	✓	✓	×	128	70.44 ± 1.30	57.74 ± 1.09	73.15 ± 0.84
12	MADDR	✓	✓	✓	✓	✓	✓	✓	128	72.38 ± 1.99	58.86 ± 0.86	75.01 ± 0.95

proposed framework, a generic network can be applied to recognize the abnormalities of all datasets. In the supplementary material, we further present the t-SNE visualization of the continuous variable part of the latent representation to show that introducing the proposed MADDR framework can deliver obviously more tightly compacted normal distributions, leading to better identification of the abnormal outliers.

3.3 Ablation Study

To further demonstrate the effectiveness of each component, we perform an ablation study with the following variants: 1) the baseline DPA method (the dimensions of z is 16) which trains networks for each organ and modality separately; 2) the baseline DPA method with the dimensions of z increased to 128; 3) using the images of LiTS, Lung-CT and Lung-X-rays datasets together to train the same DPA network; 4) imposing the organ classification constraint on the encoded latent representation z without feature disentanglement (f.d.); 5) imposing the organ classification constraint and the center constraint on the encoded latent representation z without feature disentanglement; 6) MADDR + DPA trained on LiTS and Lung-CT; 7) imposing the modality classification constraint on the encoded latent representation z without feature disentanglement; 8) imposing the modality classification constraint and the center constraint on the encoded latent representation z without feature disentanglement; 9) MADDR + DPA trained on Lung-CT and Lung-X-rays; 10) imposing both the organ and modality classification constraints on the encoded latent representation z without feature disentanglement; 11) imposing both the organ classification constraint, modality classification constraint and the center constraint on the encoded latent representation z without feature disentanglement (f.d.); 12) MADDR + DPA trained on all three datasets.

The results of all these variants are displayed in Table 2. As shown in variant 3, directly introducing more datasets for network training does not necessarily improve the performance on each dataset, due to the interference of organ and modality information. However, with the additional classification constraints, the organ and modality information can be effectively separated from the latent representation of normal distribution, such that better anomaly detection performance can be achieved for each

dataset, as shown in variants 4, 7 and 10. Furthermore, variants 5, 8 and 11 show that the center constraint can tightly compact the feature representation of normal distribution so that abnormal outliers can be easier to identify. Last but not the least, the proposed feature split strategy can help disentangle the characteristic of normal distribution from classification information, as demonstrated by variants 6, 9 and 12. For the impact of hyper-parameters, including the weights for different loss terms and the dimensions of the disentangled category representations, please refer to the supplementary material.

4 Conclusion

Unlike previous studies which committed to train exclusive networks to recognize the anomalies of specific organs and modalities separately, in this work we hypothesized that normal images of various organs and modalities could be combined and utilized to train a generic network and superior performance could be achieved for the recognition of each type of anomaly with proper methodology. With the proposed model-agnostic framework, the organ/modality classification constraint and the center constraint were imposed on the disentangled latent representation to fully utilize the available information as well as improve the compactness of representation to facilitate the identification of outliers. Four state-of-the-art methods were adopted as baseline models for thorough evaluation, and the results on various organs and modalities demonstrated the validity of our hypothesis as well as the effectiveness of the proposed framework.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (2019YFE0113900) and the National Key R&D Program of China under Grant 2020AAA0109500/2020AAA0109501.

References

1. Abati, D., Porrello, A., Calderara, S., Cucchiara, R.: Latent space autoregression for novelty detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 481–490 (2019)
2. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: semi-supervised anomaly detection via adversarial training. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11363, pp. 622–637. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20893-6_39
3. Bilic, P., et al.: The liver tumor segmentation benchmark (LiTS). arXiv preprint arXiv:1901.04056 (2019)
4. Gong, D., et al.: 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)
5. Han, C., et al.: MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC Bioinform. **22**(2), 1–20 (2021)
6. Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X., Thoma, G.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg. **4**(6), 475 (2014)

7. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
8. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
9. Perera, P., Nallapati, R., Xiang, B.: OCGAN: one-class novelty detection using GANs with constrained latent representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2898–2906 (2019)
10. Perera, P., Patel, V.M.: Learning deep features for one-class classification. IEEE Trans. Image Process. **28**(11), 5450–5463 (2019)
11. Pouyanfar, S., et al.: A survey on deep learning: algorithms, techniques, and applications. ACM Comput. Surv. **51**(5), 1–36 (2018)
12. Ruff, L., et al.: Deep one-class classification. In: International Conference on Machine Learning, pp. 4393–4402. PMLR (2018)
13. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. **54**, 30–44 (2019)
14. Shehata, M., et al.: Computer-aided diagnostic system for early detection of acute renal transplant rejection using diffusion-weighted MRI. IEEE Trans. Biomed. Eng. **66**(2), 539–552 (2018)
15. Tuluptceva, N., Bakker, B., Fedulova, I., Schulz, H., Dylov, D.V.: Anomaly detection with deep perceptual autoencoders. arXiv preprint arXiv:2006.13265 (2020)
16. Zeng, L.L., et al.: Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine **30**, 74–85 (2018)
17. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
18. Zhao, J., Zhang, Y., He, X., Xie, P.: COVID-CT-dataset: a CT scan dataset about COVID-19. arXiv preprint arXiv:2003.13865 (2020)
19. Zhou, H.Y., Chen, X., Zhang, Y., Luo, R., Wang, L., Yu, Y.: Generalized radiograph representation learning via cross-supervision between images and free-text radiology reports. Nat. Mach. Intell. **4**(1), 32–40 (2022)
20. Zhou, K., et al.: Memorizing structure-texture correspondence for image anomaly detection. IEEE Trans. Neural Netw. Learn. Syst. 2335–2349 (2021)