

Anomaly Detection for Medical Images Using Teacher–Student Model With Skip Connections and Multiscale Anomaly Consistency

Mingxuan Liu^{ID}, *Student Member, IEEE*, Yunrui Jiao^{ID}, Jingqiao Lu^{ID}, and Hong Chen^{ID}, *Senior Member, IEEE*

Abstract—Anomaly detection (AD) in medical images aims to recognize test-time abnormal inputs according to normal samples in the training set. Knowledge distillation (KD) based on the teacher–student (T-S) model is a simple and effective method to identify anomalies, yet its efficacy is constrained by the similarity between teacher and student network (S-Net) architectures. To address this problem, in this article, we propose a T-S model with skip connections (Skip-TS) which is trained by direct reverse KD (DRKD) for AD in medical images. First, to overcome the low sensitivity to anomalies caused by structural similarity, we design an encoder–decoder architecture where the teacher network (T-Net) is a pre-trained encoder and the S-Net is a randomly initialized decoder. During training, the S-Net learns to reconstruct the shallow representations of images from the output of the T-Net, which is called DRKD. Second, we introduce skip connections to the T-S model to prevent the S-Net from missing normal information of images at multiscale. In addition, we design a multiscale anomaly consistency (MAC) loss to improve the AD and localization performance. Thorough experiments conducted on 12 public medical datasets and two private medical datasets demonstrate that our approach surpasses the current state-of-the-art by 6.4% and 8.2% in terms of area under the receiver operating characteristic (AUROC) on public and private datasets, respectively. Code and organized benchmark datasets will be available at <https://github.com/Arktis2022/Skip-TS>.

Index Terms—Anomaly detection (AD), deep learning, knowledge distillation (KD), medical image analysis, teacher–student (T-S) model.

I. INTRODUCTION

IN MEDICAL image recognition, collecting and labeling abnormal data is time-consuming and expensive, especially when a disease is very rare [1]. Even with labeled data, supervised learning still faces the challenge of data imbalance,

Manuscript received 24 October 2023; revised 24 March 2024; accepted 17 April 2024. Date of publication 29 May 2024; date of current version 7 June 2024. This work was supported in part by the National Science and Technology Major Project from Minister of Science and Technology, China, under Grant 2018AAA0103100; and in part by the National Natural Science Foundation of China under Grant 92164110 and Grant 62334014. The Associate Editor coordinating the review process was Dr. Min Xia. (*Corresponding author: Hong Chen*.)

Mingxuan Liu is with the School of Biomedical Engineering, Tsinghua University, Beijing 100086, China (e-mail: liumx19@mails.tsinghua.edu.cn).

Yunrui Jiao and Hong Chen are with the School of Integrated Circuits, Tsinghua University, Beijing 100086, China (e-mail: jyr19@mails.tsinghua.edu.cn; hongchen@tsinghua.edu.cn).

Jingqiao Lu is with the XIN-Huangpu Joint Innovation Institute of Chinese Medicine, Guangzhou 510300, China (e-mail: lujingqiao@jiicm.org.cn).

Digital Object Identifier 10.1109/TIM.2024.3406792

as abnormal data is typically in lower supply compared to normal data [2]. Anomaly detection (AD) aims to recognize test-time abnormal inputs based on normal samples observed during training [3]. Because training AD models relies solely on normal data to characterize the normal distribution, and normal medical images without lesion annotation are much easier to collect [4], researchers can conveniently establish AD models. As a result, the ability of AD to accurately detect anomalies in medical images using only normal data has made it a crucial technique in artificial intelligence-assisted diagnosis.

Most state-of-the-art works realize AD by building self-supervised tasks on the training dataset, which mainly include sample reconstruction [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], pseudo-outlier augmentation [24], [25], [26], [27], and knowledge distillation (KD) [28], [29], [30], [31], [32], [33], [34]. For instance, Tschuchnig et al. [23] proposed the use of autoencoders (AEs) to quantify abnormalities in the rat spinal cord, while another study employed an anatomy-aware pasting (AnatPaste) augmentation tool [26] for AD in chest radiographs. However, both sample reconstruction and pseudo-outlier augmentation methods necessitate training models from scratch, which may lead to suboptimal performance or poor generalization. For example, Reiss et al. [35] observed that for unsupervised AD, ImageNet pre-trained features often outperform self-supervised features significantly. Therefore, feature extraction-based approaches were proposed, utilizing pre-trained encoders to directly extract image features and model their distribution for industrial AD. Noteworthy methods such as PaDiM [36] and CFA [37] have achieved state-of-the-art performance on the MVTec industrial anomaly dataset [38]. Nevertheless, the distribution of medical images is often more complex compared to natural images and exhibits larger domain deviation than the pre-training datasets, which renders the above methods less effective. The KD-based methods leverage student networks (S-Nets) to learn the distribution of pre-trained features rather than presupposing the latent distribution of features, thus is more suitable for medical images.

Previous KD-based frameworks usually use a sufficiently pre-trained teacher network (T-Net) and a S-Net with a similar or identical structure to form a teacher–student (T-S) model. During training, knowledge is transferred from the T-Net to

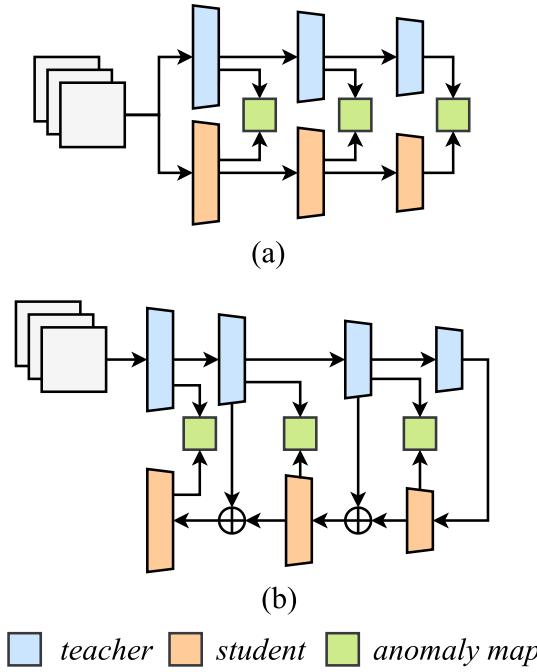


Fig. 1. T-S models and data flow in (a) traditional KD paradigm [28], [29], [30], [31] and (b) proposed DRKD paradigm. Traditional KD uses teacher and S-Nets with similar architectures, both of which receive images as input. The proposed DRKD designs the student as a decoder, and directly receives multiscale knowledge through skip connections.

the S-Net, which allows the S-Net to learn the normal data manifold. Consequently, when fed with abnormal data, the features extracted by the S-Net and T-Net are anticipated to be inconsistent. However, the structural consistency of the T-Net and S-Net in the T-S model makes it a non-distinguishing filter [30]. Specifically, both T-Net and S-Net are encoders with the same network structure. S-Net may hence tend to overfit the output of T-Net, resulting in similar internal parameters. Therefore, traditional KD-based methods cannot guarantee the large gap between the features extracted by T-Net and S-Net for abnormal images, such that it is difficult to detect subtle anomalies in medical images.

Some studies address above problems by designing a smaller S-Net [29], [30], but their anomaly localization is limited by the weak feature extraction capability of shallow networks. Deng and Li [32] proposed RDAD, a reverse distillation paradigm for the industrial AD, where a trainable one-class bottleneck embedding is utilized to connect the T-Net and S-Net, thus rendering the T-S model a framework with heterogeneity. However, this indirect connection necessitates the S-Net to recover shallow features of images from a low-dimensional representation that is dense with information. For medical tasks, normal images usually exhibit variable patterns, which will pose a challenge for the S-Net to accurately recover anomaly-free image representations and lead to AD failure consequently.

To mitigate the network structure consistency problem, in this article, we propose a novel direct reverse KD (DRKD) paradigm to train the Skip-TS model. Fig. 1 illustrates the T-S model and data flow of the proposed DRKD and the traditional KD paradigms [28], [29], [30], [31]. First, to address the

issue of non-distinguishing filters encountered in traditional KD-based methods, we propose an encoder-decoder architecture for the T-S model, where deep features extracted by the T-Net are directly passed to the S-Net without intermediaries. During training, the S-Net undergoes a process known as DRKD, whereby it learns to reconstruct the multiscale representation of the input image from its high-level semantic feature.

Second, to aid the S-Net in reconstructing an anomaly-free representation of the image, we propose a T-S model with skip connections between the T-Net and S-Net, called Skip-TS, with which the S-Net is able to receive multiscale features of the image and the reconstruction difficulty is lessened. When normal medical images are used as input, the S-Net is able to produce features that are consistent with the output of T-Net at all scale to avoid false positives. Finally, the KD-based method needs to measure the difference between each layer's output to obtain the location information of anomalies. This anomaly localization method is based on the hypothesis that feature maps at each scale can reflect anomalous features. However, this hypothesis is not always correct in practice because sometimes the feature map at a certain scale exhibits high anomaly scores in normal regions [39]. We enforce the spatial consistency of the anomaly maps across layers by adding a multiscale anomaly consistency (MAC) loss, which improves the model's ability of AD and localization.

Besides, previous studies only verified the performance of models on a small number of datasets, which is due to the lack of standardized benchmarks for medical AD. For the first time, we collect and organize 14 medical datasets, 12 of them are existing public datasets [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], and two of them are private datasets collected in our work [51]. We verify the effectiveness of our method by comparing it with current state-of-the-art methods on these datasets. In addition, we qualitatively investigate the anomaly localization performance of the proposed method.

A preliminary version of this work was previously accepted [51]. Specifically, in the conference version of our paper, the following is proposed:

- 1) We propose a novel distillation paradigm (DRKD). With this approach, the S-Net is designed as a decoder, which is directly fed with the high-level semantic features of images from the T-Net during the KD process. This effectively addresses the issue of non-distinguishing filters encountered in traditional KD-based methods.
- 2) To prevent false positives for normal medical images, a Skip-TS model is proposed to help the S-Net recover the multiscale representation of normal images by introducing skip connections between the T-Net and S-Net.
- 3) We conduct extensive experiments on five medical datasets, and the results show that our proposed AD method achieves the best performance.

In this article, we extend our earlier work [51].

- 1) We introduce a new method to accumulate the multiple anomaly maps generated by Skip-TS to realize anomaly localization.
- 2) A MAC loss is put forward to ensure the spatial consistency of the multiscale anomaly maps obtained

- by Skip-TS, which improves the AD and localization performance.
- 3) We for the first time organize 14 medical datasets (including the five datasets used in [51]) into benchmarks for medical AD. Extensive experiments on the benchmarks show that the proposed model Skip-TS has the best performance.
 - 4) We analyze the interpretability of our model by conducting t-SNE analysis on the intermediate feature maps outputted by the T-Net and S-Net of Skip-TS, to assess which image features or regions are the most important for detecting anomalies.

The rest of our paper is organized as follows. In Section II, we introduce the related work on AD. In Section III, we describe the proposed KD-based method in detail. In Section IV, extensive experiments on 14 medical datasets are conducted to verify our method. Section V concludes our work.

II. RELATED WORK

AD, also known as out-of-distribution detection, can be solved by constructing the manifold of normal data. In addition, some studies are also devoted to pixel-level AD, i.e., locating anomalies in images. This section will review previous efforts on these tasks.

A. Reconstruction-Based Methods

Recently, reconstruction-based methods have dominated medical image AD. These methods learn a mapping function to reconstruct normal samples by using generative models, such as AE [52] and generative adversarial nets (GAN) [53].

In particular, Akçay et al. [5] propose a Gandomaly model to jointly learn the generation of high-dimensional image space and the inference of latent space. The authors introduce skip connections in the model to thoroughly capture the multiscale distribution of the normal data in image space [6]. AnoGAN [7] and f-AnoGAN [8] are also GAN-based AD models, the latter trains an encoder to replace the time-consuming iterative process in the testing phase.

However, due to the lack of focus on latent feature space, GAN-based methods fail to capture all the important features of normal data, leading to incorrect detections or misclassifications. In order to overcome the shortcoming, both image and feature spaces are considered through structure similarity loss and center constraint in SALAD [4], and experiments on optical coherence tomography (OCT) and chest X-ray datasets show its effectiveness. Many approaches also use AE and its variants. Salehi et al. [11] find that introducing self-supervised tasks such as puzzle-solving into AE-based methods can prevent overfitting and facilitate learning beyond pixel-level features.

Furthermore, in order to handle complex medical images, Shvetsova et al. [12] propose deep perceptual AEs that reconstructs high-resolution images through a progressive training process. Without constraints in the training process, traditional AEs exhibit an overly generalized behavior and can sometimes recover anomalies. MAMA Net [9] utilizes a hash-addressing

memory module to solve this problem. Following the study of [9], Zhou et al. [10] propose a Proxy-bridged Image Reconstruction Network that bridges the input image and the reconstructed image with an intermediate proxy.

B. Pseudo-Outlier Augmentation-Based Methods

Pseudo-outlier augmentation converts AD to a supervised learning task by adding pseudo-anomalies in normal images. CutPaste [24] generates pseudo-outliers by cutting image patches and pasting them at random locations. To get synthetic anomaly images and reference masks for normal data, AnoSeg [25] uses hard augmentation to change the normal sample distribution. However, these methods are inefficient in medical image AD without medical expertise. AnatPaste [26] leverages a lung segmentation pretext task to generate anomalies in chest radiographs. But this technique is applied only in the AD of lung images.

C. KD-Based Methods

Some studies have found that networks pre-trained on ImageNet [54] or other large datasets can extract abnormal features, so it is possible to distinguish normal and abnormal images by directly modeling the deep features of images [36], [55], [56], [57], [58], [59], [60], [61]. Since such methods need to save a lot of features of all training samples, it results in high computational overhead. CFA [37] reduces the complexity through a scalable memory bank at the expense of the model's capacity of capturing normal data distributions, leading to inadequate AD performance on intricate datasets.

In recent years, the KD-based method has emerged as a promising solution. By designing an S-Net to transfer the feature extraction capability of the pre-trained model, and the exclusive use of normal samples for training, the T-Net and S-Net in the T-S model are capable of producing distinct features for abnormal images. In order to obtain multiscale abnormal features, US [28] uses an ensemble learning method, while STPM [29] and MKD [30] choose to calculate multiscale feature differences. Nakazawa et al. [62] further improve the model in [29] using consistency between layer groups. In RDAD [32], a reverse distillation paradigm is realized by a trainable one-class bottleneck embedding module to introduce heterogeneity in the T-S model. However, above KD-based methods are limited in medical images because of the non-distinguishing filter problem [28], [29], [30] or the difficult reconstruction task faced by the S-Net [32]. Despite attempts made by Xu et al. [33] to train an AE as a T-Net and use the KD-based method for AD in CT images, there remains a need for a generalized T-S model for medical AD.

D. Summary

In summary, most AD models for medical images use reconstruction-based schemes. However, these methods have the following problems.

- 1) The training process is complex and the performance is too sensitive to hyperparameters, which makes it hard to find the optimal settings for different datasets and tasks.

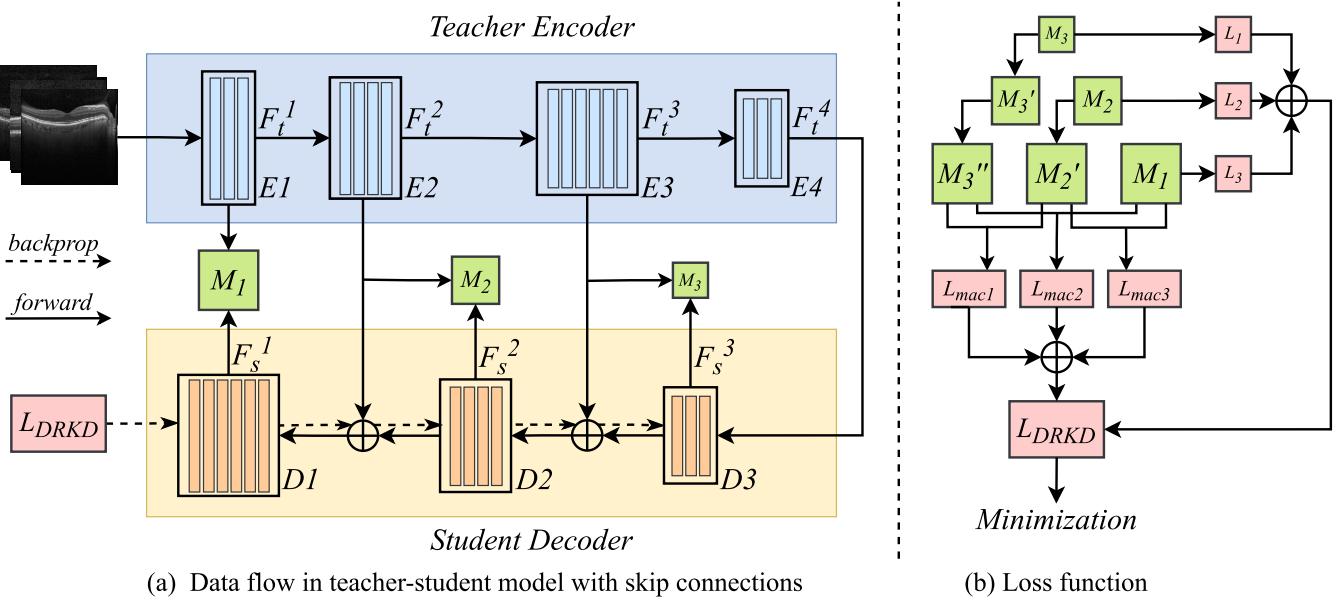


Fig. 2. Overview of the Skip-TS model and MAC loss for AD in medical images. (a) Skip-TS consists of a pre-trained T-Net and a randomly initialized S-Net. The T-Net can be divided into four sub-encoders $E1, E2, E3$, and $E4$ and the S-Net can be divided into three sub-decoders $D1, D2$, and $D3$. During training, the S-Net needs to recover the multiscale representation by minimizing the loss function L_{DRKD} . Solid lines represent the forward propagation during training and inference, while dashed lines indicate the backpropagation during training. Given that we have frozen the T-Net, the dashed lines are connected solely to the S-Net. (b) Loss function L_{DRKD} is composed of two parts, one (L_1, L_2, L_3) is directly obtained from 2-D anomaly maps $M1, M2$, and $M3$, and the other ($L_{mac1}, L_{mac2}, L_{mac3}$) comes from the difference between $M1, M2$, and $M3$.

- 2) The generative model can reconstruct both normal and abnormal regions due to its high generalization capacity, and it is difficult for the model to reconstruct the high-frequency boundaries of normal images, resulting in false positives and negatives.
- 3) Anomalies are found only by computing the pixel-level difference between the input and reconstructed images, resulting in more noise when anomaly localization.
- 4) The encoder-decoder architectures are trained from scratch, so that the powerful feature extraction capabilities of pre-trained models are not exploited, which results in less robust models and lower performance. Although previous studies have investigated the use of pseudo-outlier augmentation or KD-based methods for specific medical datasets [26], [33], these approaches suffer from limited generalization and only perform well on a few specific datasets. As such, this study aims to propose a simple, powerful, and general KD-based method that can be applied to various AD tasks in medical images.

III. METHOD

A. Pipeline

First of all, we define the AD task. Let $X_{\text{train}} = \{x_1, \dots, x_n\}$ be a training dataset containing only normal images and $Y_{\text{test}} = \{y_1, \dots, y_m\}$ be a testing dataset consisting of normal and abnormal data. Our goal is to train a model using X_{train} that can recognize and localize abnormal images in Y_{test} .

Fig. 2 illustrates the overview of the proposed Skip-TS model and MAC loss for AD in medical images. As shown in Fig. 2(a), we train a T-S model called Skip-TS through DRKD (discussed in Section III-B). The Skip-TS consists of a pre-trained T-Net and a randomly initialized S-Net. First, we adopt

a sufficiently pre-trained network to extract discriminative features of images [55]. Let F_t be the image feature extracted by T-Net, which contains normal and abnormal information. During training, the S-Net reconstructs shallow features of each image from the input. Since only normal data is included in X_{train} , the S-Net learns the patterns of normal images from F_t . Let F_s be the image feature recovered by S-Net, in the testing process, because the S-Net only learned the manifold of normal data, F_s contains only normal feature and ignores anomalies, which brings the difference between F_t and F_s for abnormal data. That is, when all the input is normal image, F_t and F_s should be highly similar. In addition, the skip connections are helpful to maintain the consistent representation of normal features.

For each y_i in Y_{test} , we can obtain the anomaly score by measuring the similarity between F_s and F_t . The lower the similarity, the higher the anomaly score. By setting a threshold with the method proposed in [63], Y_{test} can be divided into two categories: the normal subset and the abnormal subset. As a result, AD is realized. Finally, by accumulating the anomaly maps M obtained from each F_s and F_t , the location information of the anomalies can be obtained.

It should be noted that the proposed DRKD paradigm also uses an encoder-decoder architecture. But unlike the generative model, we freeze the parameters of the encoder during training. Furthermore, we pay more attention to the gap of intermediate features extracted by the T-Net and S-Net rather than the pixel difference between the input image and the reconstructed image.

B. Direct Reverse KD

As outlined in Sections I and II-C, the traditional KD paradigm lacks sensitivity toward abnormal data and the

RDAD model [32] has demonstrated a tendency to misclassify normal images as anomalies. These problems result in a low area under the receiver operating characteristic (AUROC) [64] for KD-based AD methods when applied to medical images. In order to address the problems, we put forward a DRKD paradigm shown in Fig. 2(a), from which we can see that distillation is performed on Skip-TS with an encoder-decoder architecture, and features extracted by the T-Net are directly transferred to the S-Net without any intermediary.

With DRKD, T-Net is used to extract comprehensive features from input images including normal and abnormal information. For instance, when utilizing the pre-trained WideResNet50 [65] on ImageNet [54] and assigning $E1$, $E2$, $E3$, and $E4$ to its four layers. In this configuration, $E4$ generates the input for S-Net, which represents the low-resolution feature set of the images. The task of other three sub-encoders $E1$, $E2$, and $E3$ is to provide the multiscale knowledge as a reference for the S-Net. During training, the parameters of T-Net are frozen to prevent the model from converging to a trivial solution [31].

The S-Net is structured symmetrically to that of T-Net to learn the intermediate representation of normal images from T-Net. This symmetry ensures that the output F_s of each layer in S-Net maintains consistent dimensions with its corresponding layer output F_t in T-Net. As a result, this design effectively prevents S-Net from processing anomaly data directly during the testing phase. In particular, when using WideResNet50 as T-Net, downsampling is accomplished through convolutional layers with a kernel size of 3 and a stride of 2 [65]. Accordingly, the S-Net employs deconvolution layers [66] with a kernel size of 3 and a stride of 2 for upsampling.

For the input $x_i \in R^{w \times h \times c}$ where h is the height, w is the width, and c is the number of color channels, the k th sub-encoder of the T-Net outputs $F_t^k(x_i) \in R^{w_k \times h_k \times d_k}$, where w_k , h_k and d_k represent the width, height, and channel number of the feature map, respectively. It is assumed that the number of sub-encoders is a , and $(a-1)$ for sub-decoders, the $(a-1)$ th sub-decoder takes $F_t^a(x_i)$ as input and outputs $F_s^{a-1}(x_i) \in R^{w_{a-1} \times h_{a-1} \times d_{a-1}}$. For j th ($j \in (0, a-1)$) sub-decoder, the input becomes $(F_t^{j+1}(x_i) + F_s^{j+1}(x_i))$ because of the existence of skip connections (discussed in Section III-C), and its output is $F_s^j(x_i) \in R^{w_j \times h_j \times d_j}$. It should be noted that when k is equal to j , that is, the dimension of $F_t^k(x_i)$ is equal to that of $F_s^j(x_i)$. We adopt the cosine similarity between F_t and F_s to compute the anomaly map. For F_t^l and F_s^l ($l \in (0, a)$), we can obtain a 2-D anomaly map $M_l(x_i) \in R^{w_l \times h_l}$ by calculating the cosine similarity along the channel axis

$$M_l(w, h) = 1 - \frac{(F_t^l(w, h))^T \cdot F_s^l(w, h)}{\|F_t^l(w, h)\| \cdot \|F_s^l(w, h)\|}. \quad (1)$$

1) Similarity Loss: As shown in Fig. 2(b), the output features of the T-Net and S-Net are required to be consistent when inputting a normal image to the model, that is, each pixel of the anomaly map should be close to zero. Therefore, we can constrain the output of S-Net by setting a similarity loss. The loss function L_l of each layer can be obtained by

accumulating the loss map

$$L_l = \frac{1}{w_l h_l} \sum_{w=1}^{w_l} \sum_{h=1}^{h_l} M_l(w, h) \quad (2)$$

then, the loss functions of each layer are added together to obtain the similarity loss L_{sim}

$$L_{\text{sim}} = \sum_{l=1}^{a-1} L_l. \quad (3)$$

2) MAC Loss: When KD-based methods are used for anomaly localization, it is commonly assumed that anomalous characteristics can be reflected across various scales by anomaly maps. However, this assumption may fail in certain scenarios. Inconsistencies in anomaly locations depicted by the anomaly map on a particular scale can undermine the final detection and localization of anomalies. In order to ensure that each scale's anomaly map accurately identifies anomalies, we use the MAC loss function to constrain the anomaly maps produced by the layers of the T-S model. First, we upsample each anomaly map to the same scale as the largest anomaly map via bilinear interpolation and get $M_i^{\text{bi}}(x_i) \in R^{w_{a-1} \times h_{a-1}}$. After that, we calculate the difference between all maps to obtain the MAC loss L_{mac}

$$L_{\text{mac}} = \frac{1}{w_{a-1} h_{a-1}} \times \sum_{w=1}^{w_{a-1}} \sum_{h=1}^{h_{a-1}} \sum_{i,j \in (0,a)} (M_i^{\text{bi}}(w, h) - M_j^{\text{bi}}(w, h))^2. \quad (4)$$

Finally, with the loss function defined above, the overall objective of optimizing the student decoder can be expressed as

$$L_{\text{DRKD}} = L_{\text{sim}} + \beta \cdot L_{\text{mac}} \quad (5)$$

where β is the loss weight and it is set to 0.05.

C. Skip Connections

In order to help the S-Net reconstruct anomaly-free representations of images, we introduce skip connections to skip-TS inspired by image reconstruction [67]. For j th ($j \in (0, a-1)$) sub-decoder, we let the $(F_t^{j+1}(x_i) + F_s^{j+1}(x_i))$ as its input. For abnormal images, as the S-Net is trained with normal data, $F_s^{j+1}(x_i)$ only contains normal patterns of images, and the sub-decoder learns how to decode the $F_s^{j+1}(x_i)$ into shallow features. However, because the pre-trained network T-Net can fully extract image features, $F_t^{j+1}(x_i)$ can be divided into $F_{t1}^{j+1}(x_i)$ and $F_{t2}^{j+1}(x_i)$

$$F_t^{j+1}(x_i) = f(F_{t1}^{j+1}(x_i), F_{t2}^{j+1}(x_i)) \quad (6)$$

where $F_{t1}^{j+1}(x_i)$ and $F_{t2}^{j+1}(x_i)$ are normal and abnormal features, respectively. The sub-decoder successfully decodes $F_{t1}^{j+1}(x_i)$ based on the knowledge gained during training but cannot decode $F_{t2}^{j+1}(x_i)$, which leads to a gap between $F_t^j(x_i)$ and $F_s^j(x_i)$. On the other hand, for normal images in testing, the information in $F_t^{j+1}(x_i)$ helps the sub-decoder reconstruct the shallow representation.

TABLE I
KEY STATISTICS FOR IMAGE DATASETS

Public Data	ID	Imaging Modalities	Associated Diseases	Training Data Normal	Testing Data Normal & Abnormal
HeadCT [40]	1	CT	Brain hemorrhage	60	40 & 100
BrainMRI [41]	2	MRI	Brain tumor	58	40 & 154
LungCT [42]	3	CT	Covid19	432	145 & 811
BUSI [43]	4	Ultrasound	Breast cancer	99	34 & 647
BreastMRI [44]	5	MRI	Breast cancer	525	175 & 700
Retinal OCT [45]	6	OCT	Diabetic macular edema	542	50 & 735
CP-CHILD-A [46]	7	Colonoscopy	Colonic polyps	800	800 & 800
CP-CHILD-B [46]	8	Colonoscopy	Colonic polyps	800	300 & 300
NAFLD [47]	9	Whole slide imaging	Non-alcoholic fatty liver	525	1645 & 2150
IQ-OTH/NCCD [48]	10	CT	Lung cancer	277	139 & 681
Covid-Xray [49]	11	Xray	Covid19	70	20 & 132
Malaria-Cell [50]	12	Microscopy imaging	Malaria	344	345 & 330
Private Data	ID	Imaging Modalities	Associated Diseases	Training Data Normal	Testing Data Normal & Abnormal
TongueNet-A [51]	13	Camera imaging	Rotten coating tongue	519	484 & 9
TongueNet-B [51]	14	Camera imaging	Pealed coating tongue	454	1363 & 30

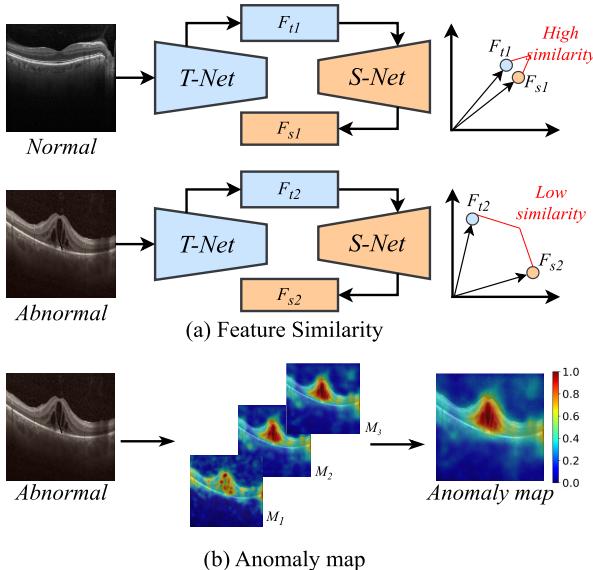


Fig. 3. Flowchart of AD and localization. (a) During testing, the T-Net outputs the real feature F_t and the S-Net outputs the anomaly-free one F_s . The anomaly score can be obtained by calculating the difference between F_t and F_s . (b) Anomaly localization can be achieved by merging anomaly maps on multiscale. Red areas indicate higher anomaly scores.

In summary, we argue that skip connections ensure consistent features when using normal images as well as detect anomalies.

D. AD and Localization

In the testing phase, we first implement anomaly localization. As shown in Fig. 3(a), KD-based methods assume a low similarity between F_t and F_s when inputting abnormal images. Based on this assumption, for image $x_i \in R^{w \times h \times c}$, we calculate the anomaly maps of each scale $M_l(x_i) \in R^{w_l \times h_l}$ through (1), and upsample all anomaly maps to the same size as the input image through bilinear interpolation, finally the final anomaly map $M_{\text{ano}}(x_i) \in R^{w \times h}$ is the element-wise accumulation of them. Fig. 3(b) shows the acquisition process

of the final anomaly map

$$M_{\text{ano}}(x_i) = \sum_{l=1}^{a-1} \text{Upsample}(M_l(x_i)). \quad (7)$$

For AD, we directly take the maximum value of $M_{\text{ano}}(x_i)$ as the anomaly score of x_i

$$S_{\text{AD}}(x_i) = \max(M_{\text{ano}}(x_i)). \quad (8)$$

IV. EXPERIMENTS

A. Datasets

Fourteen datasets including two private datasets collected in our work are adopted to evaluate the proposed method as organized benchmarks. The key information and division methods of these datasets are shown in Table I. It should be noted that the training dataset only contains normal images, while the testing dataset consists of both normal and abnormal images. And the abnormal images are provided by patients with associated diseases shown in Table I. To illustrate the distribution of anomalies across the datasets, we generate t-SNE embeddings for both normal and abnormal images, as shown in Fig. 4. Additionally, to facilitate subsequent comparisons, we assign each dataset with an ID.

The difference between our organized AD benchmarks and the datasets used in previous works are as follows.

- 1) *More dataset:* Most previous works verified their model on less than five datasets. For example, an OCT dataset and an Xray dataset are used in SALAD [4]. MAMA Net [9] and ProxyAno [10] were compared with other models on three datasets. The benchmark published by Cai et al. [73] only includes three CXR datasets, one brain MRI dataset, and one retina fundus image dataset. As we know, more datasets with different imaging modalities will help to fully test the generality of the model. Therefore, we propose the organized benchmarks with 14 datasets encompassing various imaging modalities. According to the classification method proposed by Tschuchnig and Gadermayer [74], we classify the datasets into four categories: ocular medical images (like Retinal OCT dataset), medical images of the central body region

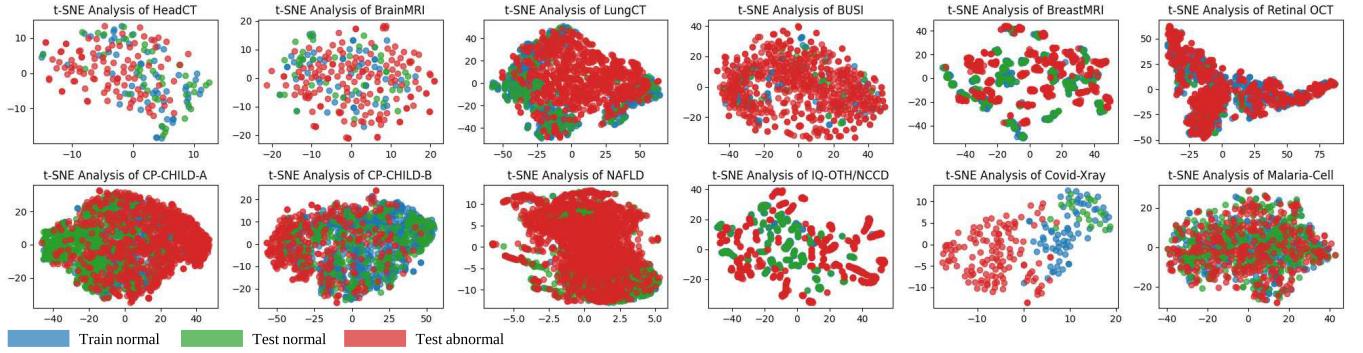


Fig. 4. t-SNE embeddings of train normal images (blue), test normal images (green), and test abnormal images (red) from the 12 public datasets.

(like LungCT, BUSI, BreastMRI, CP-CHILD-A, CP-CHILD-B, IQ-OTH/NCCD, and Covid-Xray datasets), brain images (such as HeadCT and BrainMRI datasets), and tissue or cellular images (such as NAFLD and Malaria-Cell datasets). In addition, we collect two unique datasets of tongue images obtained via camera and have pioneered the application of AD methods to these tongue image datasets.

- 2) *Small sample size*: Due to the high cost of collecting medical data, the medical dataset usually includes less than one thousand medical samples. To adapt to the situation, we set the number of samples in the training dataset to less than a thousand. In addition, as can be seen from Table I, for some datasets (HeadCT, BrainMRI, BUSI, and Covid-Xray), no more than 100 images are used as training data in our work to obtain the AD performance of models with such very small samples.
- 3) *Dedicated datasets for AD*: In real application, AD models are usually used to find abnormal images when there are too few abnormal images to train supervised models. Therefore, we collect two private datasets (TongueNet-A and TongueNet-B) to demonstrate such scenarios. The tongue images in the datasets were collected from several traditional Chinese medicine (TCM) hospitals and labeled by 30 senior doctors. According to the pathological characteristics of tongue coating, the abnormal tongue images in TongueNet-A have rotten coating while peeled coating for TongueNet-B. It can be seen from Table I that our private datasets contain only 9 and 30 abnormal images, respectively, which makes it impossible to train the supervised model to achieve binary classification, and only AD models are able to detect abnormal images.

B. Implementation Details

All images in datasets are uniformly resized to 256×256 . We use Adam optimizer [75] with the learning rate of 0.005 and $\beta = (0.5, 0.999)$. The training epoch is set 200, and batch size is 16. For the T-Net, we use the pre-trained WideResNet50, and the architecture of S-Net is symmetrical to that of the WideResNet50. Note that we only train the S-Net and freeze the T-Net. All experiments are implemented using PyTorch and conducted on the NVIDIA GeForce RTX 3090.

C. Evaluation Metrics

Considering that the choice of threshold will greatly affect indicators such as F1-score [10], we adopt AUROC [64] to evaluate AD performance. The AUROC measures how well the model can separate normal samples from abnormal ones. In order to verify the robustness of our models, we test the last epoch under five different random number seeds as the final result for all models. Note that this is a very harsh experimental setup and we do not allow any unprincipled early stops [11], [30].

D. Competing Methods

We compare our method with several state-of-the-art AD methods, including one-class classification model FCDD [68], reconstruction-based PuzzleAE [11], f-AnoGAN [8], Ganomaly [5], pseudo-outlier augmentation-based CutPaste [24], feature extraction-based CFA [37], PaDiM [36], OrthoAD [69], and Patchcore [56], KD-based RDAD [32], STPM [29], and DFC [70]. In addition, we also test two state-of-the-art open-set supervised models for AD, DevNet [71] and DRA [72], to observe the limitations of our proposed models. Such open-set supervised models need to use a few labeled abnormal images as examples. In the experiments, we use the code provided in the GitHub to implement the above methods. In training DevNet and DRA, we use training dataset and one abnormal image.

E. AD Results

We report the AUROC (%) of our method and competing methods on 14 datasets in Table II, and highlight the best four results on each dataset. It can be observed that for the average outcome our method exceeds state-of-the-art by **6.4%** ($86.1\% \rightarrow 92.5\%$) on public datasets, and **8.2%** ($80.2\% \rightarrow 89.0\%$) on private datasets. However, the dataset has wide range of sample size hence simply averaging over AUROC percentages could lead to biased result. To solve the problem, we also calculate the weighted AUROC, where the size of each dataset is used as a weighting factor during the averaging process. The results, as shown in Fig. 5(b), demonstrate that Skip-TS is the best-performing model, achieving weighted AUROC of **92.7%** and **86.2%** on the public and private datasets, respectively. In addition, in order to visualize the misclassified and correctly classified classes, we also generated

TABLE II
AD AUROC (%) ON CONSIDERED DATASETS (THE BEST FOUR RESULTS ARE SHOWN IN **BLUE**, **GREEN**, AND **BLACK** FONT)

Methods	Dataset ID															
	1	2	3	4	5	6	7	8	9	10	11	12	AVG	13	14	AVG
FCDD [68]	81.1	59.4	65.5	67.5	70.8	78.9	79.6	79.0	82.2	92.2	92.1	83.8	77.7	81.3	64.1	72.7
Ganomaly [5]	74.0	73.8	58.6	68.0	73.2	64.2	75.2	92.8	71.2	98.7	30.5	57.9	69.8	70.8	45.7	58.3
f-AnoGAN [8]	82.6	69.5	89.1	77.1	90.4	70.7	97.0	95.5	83.8	89.2	93.8	67.3	83.8	83.6	63.3	73.5
PuzzleAE [11]	83.8	71.1	75.2	81.8	89.9	75.1	92.7	96.4	79.9	98.9	69.7	63.5	81.5	81.9	55.9	68.9
CutPaste [24]	73.0	67.0	83.3	79.4	86.7	93.9	89.6	87.4	88.4	95.8	94.7	71.3	84.2	72.4	66.7	69.6
OrthoAD [69]	70.4	82.1	71.2	88.0	62.4	89.6	76.6	63.3	85.6	98.1	78.3	73.8	78.3	77.1	56.0	66.6
PaDiM [36]	77.1	85.3	77.0	93.8	80.4	96.3	85.0	80.9	79.5	98.1	88.2	87.4	85.8	73.6	66.1	69.9
CFA [37]	68.8	76.1	71.2	85.6	58.2	67.8	75.0	64.2	74.9	82.0	79.1	59.6	71.9	53.8	54.1	54.0
Patchcore [56]	72.4	65.5	80.5	81.5	72.0	94.5	97.8	96.5	91.1	84.3	78.5	84.8	83.3	80.8	70.3	75.6
STPM [29]	75.7	77.9	79.0	82.5	73.5	62.9	82.0	85.9	85.1	94.7	90.6	94.0	82.0	87.0	61.2	74.1
DFC [70]	65.6	70.2	74.2	91.7	95.9	88.5	92.8	92.0	87.9	93.8	83.3	97.5	86.1	72.9	58.1	65.5
RDAD [32]	74.3	80.9	89.3	88.6	85.9	88.5	89.1	80.5	89.4	95.6	88.2	83.3	86.1	92.4	69.2	80.8
DevNet [71]	84.2	63.3	92.1	68.4	55.9	84.8	97.6	82.1	89.8	90.0	98.9	92.1	83.3	80.1	63.3	71.7
DRA [72]	83.8	67.4	90.9	69.6	54.4	85.2	98.1	85.4	92.2	83.2	97.5	71.7	81.6	85.6	82.3	84.0
Ours	85.7	88.2	90.4	92.5	93.5	91.3	96.5	95.3	90.3	97.3	94.1	95.2	92.5	94.8	83.1	89.0

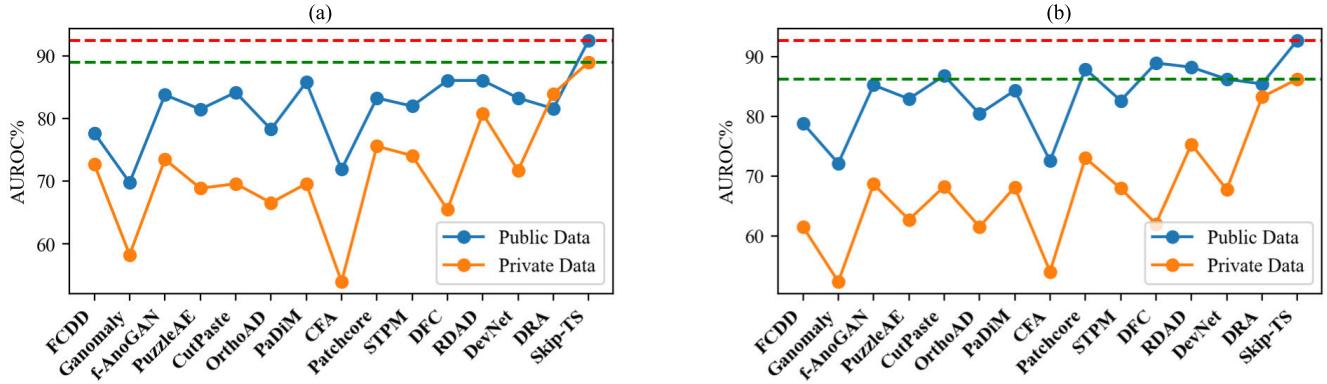


Fig. 5. AD (a) AUROC% and (b) weighted AUROC% comparison of different models.

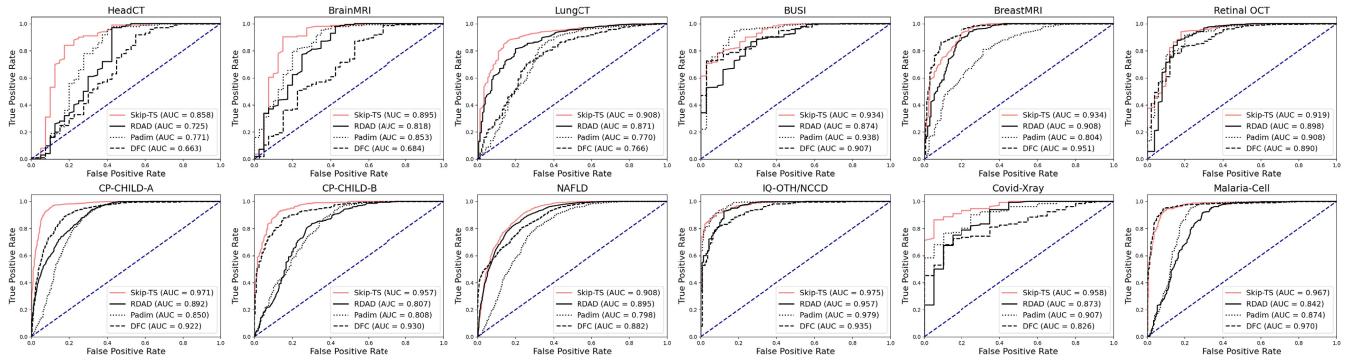


Fig. 6. AUROC curves of the top four performing models (Skip-TS, RDAD, DFC, Padim) on public datasets.

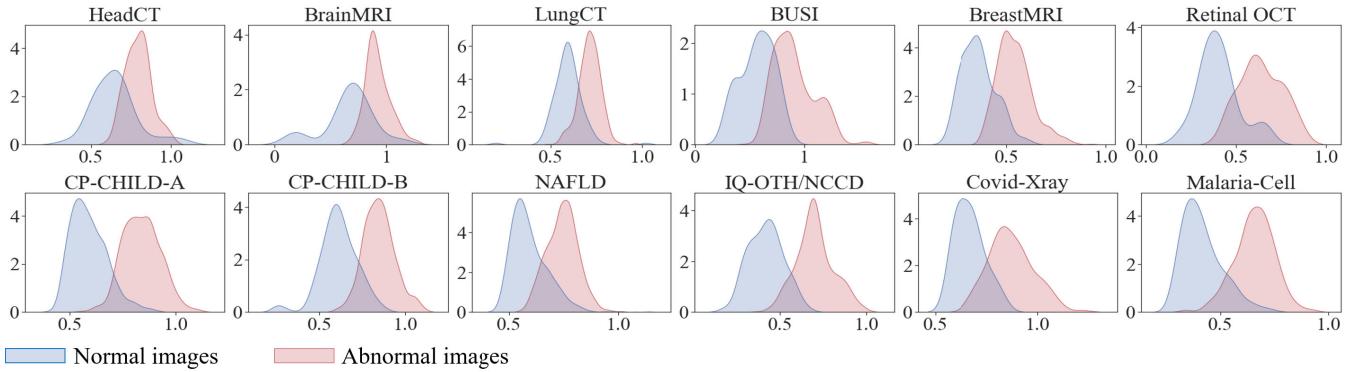
the AUROC curves for the top four ranked models on the public dataset, as depicted in Fig. 6. Finally, the optimal threshold for each dataset is determined by maximizing the difference between the true positive rate (TPR) and the false positive rate (FPR). The confusion matrices are illustrated in Fig. 7, from which we find that across 12 public datasets, Skip-TS achieves F_1 scores above 0.85 consistently, indicating that our model can attain high classification accuracy after being

trained solely on normal data. From further analysis of the data in Table II, we can get the following conclusions.

1) *Generalization:* Excluding the open-set supervised models DevNet and DRA, our Skip-TS emerges as the sole model that consistently ranks top four and achieves an AUROC of over 80% on all datasets. This demonstrates that the proposed method has robust generalization and works effectively on datasets with different imaging modalities, whereas other

		HeadCT		BrainMRI		LungCT		BUSI		BreastMRI		Retinal OCT	
		Accuracy: 0.84	F1 Score: 0.88	Accuracy: 0.89	F1 Score: 0.93	Accuracy: 0.87	F1 Score: 0.92	Accuracy: 0.79	F1 Score: 0.88	Accuracy: 0.91	F1 Score: 0.94	Accuracy: 0.93	F1 Score: 0.96
True label	Normal	33	7	34	6	122	23	31	3	140	35	41	9
	Anomaly	16	84	15	139	104	707	139	508	48	652	44	691
		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label	
True label	Normal	CP-CHILD-A		CP-CHILD-B		NAFLD		IQ-OTH/NCCD		Covid-Xray		Malaria-Cell	
	Anomaly	711	89	263	37	1241	404	128	11	19	1	310	35
		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label	
True label	Normal	28	772	25	275	185	1965	60	621	18	114	19	311
	Anomaly	711	89	263	37	1241	404	128	11	19	1	310	35
		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label		Predicted label	

Fig. 7. Confusion matrices at the optimal threshold for the proposed Skip-TS model on public datasets.

Fig. 8. KDE plots of anomaly scores for all considered public datasets by our method (x -axis: anomaly score S_{AAD} and y -axis: density).

methods are only applicable on a subset of datasets. This is attributed to the thoroughly pre-trained backbone network of the T-Net in Skip-TS model, enabling the extraction of key features from various image types. Additionally, we design the inverse S-Net to address the non-distinguishing filter issue, taking advantage of both AE-based and KD-based methods.

2) *Small-Sample Learning*: For HeadCT, BrainMRI, BUSI, and Covid-Xray with IDs of 1, 2, 4, and 11, respectively, less than 100 images are used for training. For HeadCT and BrainMRI, our method reaches new state-of-the-art AUROC of **85.7%** and **88.2%**, respectively. In addition, with our method, the AUROC is only **1.3%** (93.8% \rightarrow 92.5%) lower than that of PaDiM on BUSI and only **0.6%** (94.7% \rightarrow 94.1%) lower than that of CutPaste on Covid-Xray. Therefore, the proposed method has the best performance overall on the small-sample dataset.

3) *Pre-Trained Models*: Upon analyzing the average results, we find that the top four performing methods use pre-trained models. Notably, for the public datasets, our proposed method ranks first, followed by RDAD, DFC, and PaDiM,

respectively. Similarly, for the private datasets, the order of the top-performing methods is Ours, RDAD, Patchcore, and STPM. This outcome is largely attributed to the efficacy of pre-trained models in feature extraction, a trait which our method leverages to its fullest potential. It is worth noting, however, that most AD methods for medical images do not employ pre-trained models but instead prioritize training generative models from scratch [4], [7], [8], [9], [10], [12].

4) *Open-Set Supervised Models*: Open-set supervised models utilize a limited number of abnormal images as a training set to enhance performance. Table II indicates that DevNet outperforms all unsupervised models on LungCT and Covid-Xray, while DRA achieves the state-of-the-art on CP-CHILD-A and NAFLD. Nevertheless, the average performance of DevNet and DRA on public datasets is inadequate, especially with regards to the AUROC on BreastMRI, which is remarkably low (55.9% and 54.4%, respectively). This is due to their susceptibility to overfitting to visible anomalies, resulting in low robustness, which makes them unsuitable for datasets with variable anomaly images. Therefore, unsupervised models remain the best option for medical AD.

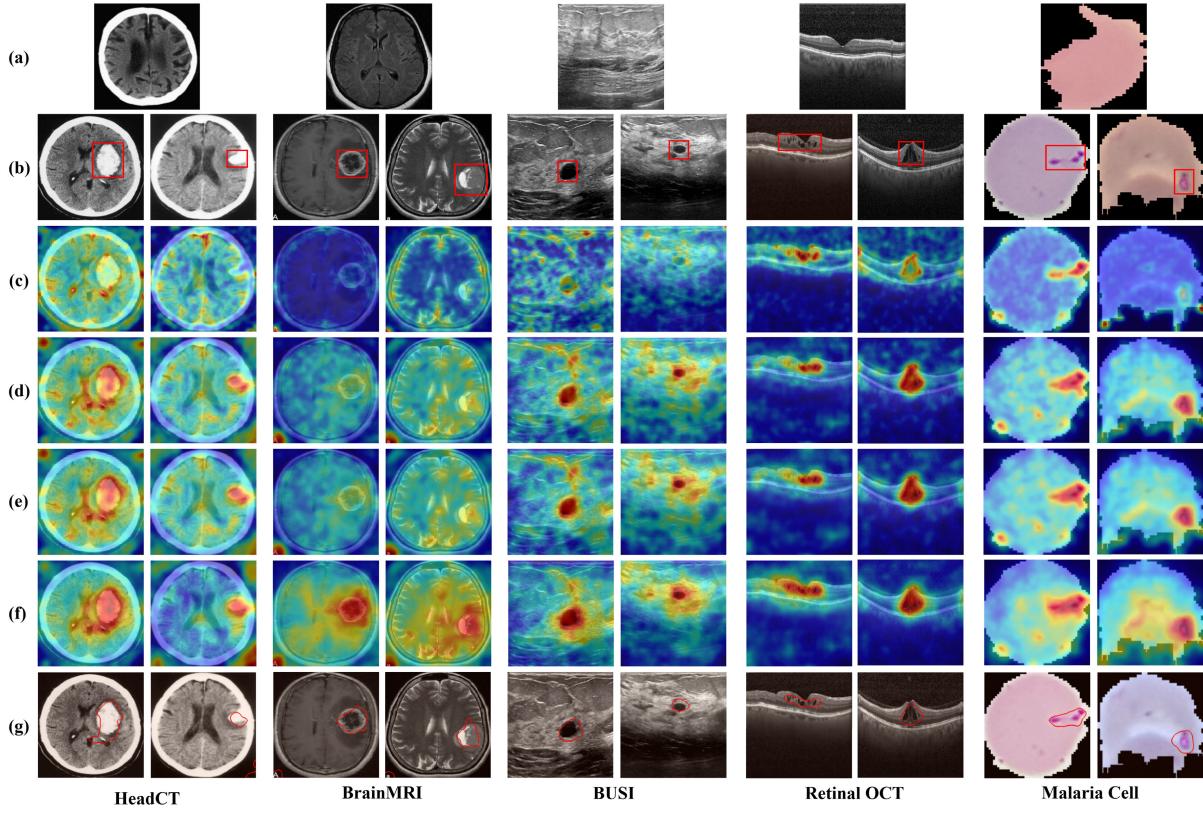


Fig. 9. Anomaly localization results of the proposed method. Our model can precisely localize anomalies, whereas most reconstruction-based methods can only roughly detect them. (a) Normal images. (b) Abnormal images (the red rectangle denotes the abnormal area). (c) Anomaly map M_1 . (d) Anomaly map M_2 . (e) Anomaly map M_3 . (f) Final anomaly map M_{ano} . (g) Segmentation results.

5) *Dataset Types*: In histology, it is insufficient to detect large objects (such as tumors) merely by their different intensity values. Besides the intensity values, the shape and interaction of nuclei and cells include high-frequency information [74]. Consequently, methods based on reconstruction struggle to differentiate between normal and anomalous data within such datasets. Within the 14 datasets, NAFLD is a histological set, and Malaria-Cell is cellular. For the two datasets, reconstruction-based methods underperform compared to KD-based methods. For instance, PuzzleAE's AUROC on NAFLD and Malaria-Cell is 79.0 and 63.5, respectively, which is 10.4 and 31.7 points lower than Skip-TS, which scores 90.3 and 95.2, respectively, on the two datasets.

6) *Kernel Density Estimate*: Finally, kernel density estimate (KDE) plots of anomaly scores S_{AD} of our method are shown in Fig. 8. We can intuitively see that the S_{AD} of normal (blue) and abnormal (red) images have non-overlapping distributions, which indicates that the proposed method has good AD ability.

F. Anomaly Localization Results

Noting that the anomalies in some datasets concentrate in local regions, we test the anomaly localization ability of the proposed method on these datasets. As shown in Fig. 9, we select five datasets, HeadCT, BrainMRI, BUSI, Retinal OCT, and Malaria Cell, to show the segmentation results in each step. The original medical images are shown in Fig. 9(a) and the abnormal images marked with red rectangles

TABLE III
COMPUTATIONAL COSTS OF SKIP-TS AND RDAD

Model	Computational Complex	
	FLOPs	Parameters
RDAD	3.72E+10	1.93E+08
Skip-TS	2.98E+10	1.33E+08

are shown in Fig. 9(b). First, we obtain the anomaly maps M_1 , M_2 , and M_3 by calculating the feature differences of each hierarchy (shown in Fig. 9(c)–(e), respectively). Second, we get the final anomaly map M_{ano} [shown in Fig. 9(f)] according to (7) and Gaussian filtering with $\sigma = 4$. Setting the threshold as 0.8, the segmentation results are shown in Fig. 9(g). Different from reconstruction-based methods that can only roughly locate anomalies by computing pixel-level differences [4], [8], [9], [10], our proposed method enables precise recognition and localization of anomalies on diverse images [shown in Fig. 9(g)], which are similar to those obtained by the semantic segmentation model [76] without using abnormal image and mask for training. Moreover, our analysis reveals that the segmentation outputs produced by our method fail to fully capture the abnormal regions in the HeadCT and BrainMRI datasets, implying that the performance of location may be hindered by the limited availability of training data. However, existing studies often ignore the impact of limited data.

TABLE IV
ABLATION STUDY OF THE PROPOSED METHOD (SC: SKIP CONNECTIONS)

	Public datasets	Private datasets
Distillation paradigm and skip connections		
Traditional KD	82.6	76.2
DRKD w/o SC	88.0	73.6
Loss functions		
L_{sim} (<i>MSE</i>)	92.4	86.9
L_{sim} (<i>MSE</i>) + L_{mac}	92.2	86.7
L_{sim} (<i>Cosine</i>)	92.0	87.8
Multi-scale feature		
$M1+M2$	86.4	78.3
$M1+M3$	91.4	87.6
$M2+M3$	92.5	87.8
Our method		
DRKD & SC & (L_{sim} (<i>Cosine</i>)) + L_{mac}) & M_{ano}	92.5	89.0

G. Computational Costs and Parameter Quantities

We compare the computational costs and parameter quantities of our proposed model, Skip-TS, with the state-of-the-art model, RDAD. The experimental results are presented in Table III, where it can be observed that our model exhibits superior computational efficiency. Our model requires only **2.98E + 10** FLOPs, which is 20% fewer than the FLOPs required by RDAD (3.72E + 10 FLOPs). Additionally, our model has 31% fewer parameters, with a count of **1.33E + 08**, compared to RDAD's parameter count of 1.93E + 08. This difference arises from our use of simple skip connections to transfer multiscale features extracted by the T-Net, instead of employing a trainable one-class bottleneck embedding module.

H. Interpretability Analysis

To analyze which image features or regions are the most important for detecting anomalies, we choose CP-CHILD-B [46] dataset as an example. Since three scale features extracted by T-Net and S-Net, respectively, are used for the computation of three anomaly maps, $M1$, $M2$, and $M3$, we conduct t-SNE analysis on the multiscale features F_t^1 , F_t^2 , and F_t^3 extracted by T-Net, as well as the features F_s^1 , F_s^2 , and F_s^3 extracted by S-Net. The analysis results are shown in Fig. 10, from which we can see that for T-Net, the extracted features F_t^1 , F_t^2 , and F_t^3 exhibit a significant overlap between the normal and anomalous data after t-SNE dimensionality reduction. This is because although pre-trained models are good at extracting features from images, they are unable to specifically highlight the anomalous features of images, especially in medical images that differ significantly from natural images. Therefore, models directly learning the feature distribution extracted by pre-trained models, i.e., feature-based models (orthoAD [69], PaDiM [36], CFA [37], and Patchcore [56]), demonstrate suboptimal performance in medical image AD. However, for S-Net trained only on normal data, normal and abnormal images can be clearly distinguished as illustrated in the first row in Fig. 10.

We further evaluate the AUROC (%) of our model using anomaly maps with different scales. When using only $M1$ obtained from F_s^1 and F_t^1 , the AUROC (%) is 63.8, whereas when using only $M2$ and $M3$, it is 85.7 and 95.1,

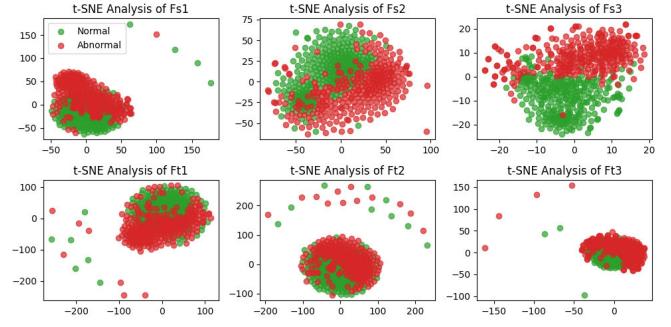


Fig. 10. t-SNE embeddings of multiscale features $\{F_t^1, F_t^2, F_t^3\}$ and $\{F_s^1, F_s^2, F_s^3\}$ extracted from the CP-CHILD-B dataset by T-Net and S-Net.

respectively. When combining $M1$, $M2$, and $M3$, the AUROC (%) is 95.7. Therefore, the importance of AD across the three scales is ranked as follows: $M3 > M2 > M1$. This aligns with our ablation experimental results, the AUROC (%) ranking is $M2 + M3 > M1 + M3 > M1 + M2$ when calculating anomaly scores using various anomaly map combinations (shown in Table IV). In the work with similar interpretability analysis, Wang et al. [29] studied the impact of various layer features extracted by ResNet18 on AD when using STPM, the results indicate that deeper features enhance the performance of AUROC. One of the reasons that reconstruction-based methods exclusively focus on pixel-level discrepancies between input and reconstructed images, without deeper feature levels. Although the SALAD model, introduced by Zhao et al. [4], accounts for differences in both image and feature spaces, its inability to compute multiscale anomaly maps precludes precise localization of anomalous regions.

In summary, features extracted by the S-Net are more discriminative between normal and anomalous instances compared to those from the T-Net, resulting in better AD performance of KD-based methods over feature extraction-based methods. With the more distinctive nature of deeper feature differences, KD-based approaches outperform reconstruction-based methods. For Skip-TS, the depth features F_t^3 and F_s^3 are pivotal for AD, with the fusion of three-scale features yielding the optimal performance.

I. Ablation Study

After achieving state-of-the-art performance on the organized benchmark, we conduct an ablation study with the proposed method in terms of distillation paradigm, skip connections, loss functions, and multiscale features. We delete one component each time and keep the other parts unchanged to test the function of each module. The results are shown in Table IV.

1) DRKD and SC for AD: To begin, we verify the effectiveness of DRKD and skip connections in improving the performance of our model. From Table II, we can find that the performance of our model with DRKD comprehensively surpasses the STPM with the traditional KD paradigm. For more fair comparison, we redesign a T-S model with our network structure, and train the model with the traditional KD paradigm in STPM. Besides, both T-Net and S-Net are set as encoders with WideResNet50 as backbone, and other experimental settings remain the same. As can be seen from

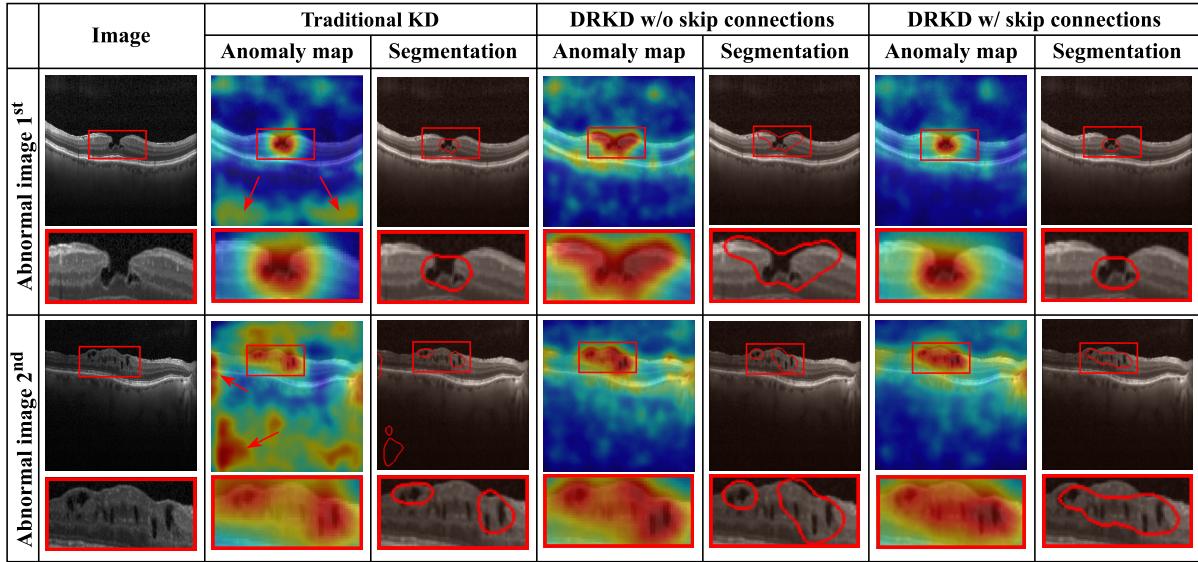


Fig. 11. Anomaly localization results with different distillation paradigms. Within each row, the sub-row below shows a zoomed-in view of the image above.

Table IV, compared with traditional KD, DRKD paradigm improves AUROC by **5.4%** (82.6% → 88.0%) on public datasets, but slightly decreases on the private dataset (76.2% → 73.6%). Skip connections further extend the advantages of DRKD, increasing the AUROC on public and private datasets to **92.5%** and **89.0%**, respectively, achieving state-of-the-art performance.

2) *DRKD and SC for Anomaly Localization*: We qualitatively investigate effect of DRKD and skip connections for anomaly localization, and the results are shown in Fig. 11. As discussed in Sections III-B and III-C, the traditional KD paradigm is not sensitive enough to anomalies and DRKD without skip connections may lead to false identification of normal regions as anomalies. We can see these phenomena in Fig. 11. For traditional KD, we first pay attention to the anomaly map, from which we find that there are a large number of high-heat areas (pointed by the arrow) outside the anomaly area. This is because the structural similarity of the T-S model in traditional KD paradigm makes it non-discriminative for anomalies. Although the segmentation result for the first anomaly image is good, it cannot accurately locate the second anomaly image and is accompanied by noise. For DRKD without skip connections, due to the difficulty of the S-Net to adequately reconstruct normal regions, the segmentation range is too large for the first abnormal image. For the second one, although the false positive area is reduced compared with the traditional KD, all abnormal areas are still not detected. However, after adding skip connections, the T-S model outputs accurate segmentation results for both abnormal images. Particularly with the first input anomaly image, when skip connections are not employed, erroneously high anomaly scores are produced for the regions adjacent to the anomaly. However, after incorporating skip connections, the S-Net gains access to multiscale features of the image, which eases the reconstruction challenge, allowing normal areas to be accurately reconstructed with correspondingly lower anomaly scores.

3) *Loss Function*: We compare four loss function selection strategies. For similarity loss L_{sim} , besides the cosine similarity loss function, the mean squared error (MSE) loss function is also commonly used. From Table IV, we can see that the combination of $L_{\text{sim}}(\text{Cosine})$ and MAC loss L_{mac} , i.e., our method achieves the best performance.

4) *Multiscale Feature*: Since the feature maps from each layer of the pre-trained network can identify abnormalities of different scales [55], in Section III-D, we advocate accumulating all anomaly maps M_1, M_2, M_3 to calculate the abnormal score of each image. In this section, we explore the influence of multihierarchical features. From Table IV, we see that removing M_1 brings the least impact on overall performance, reducing AUROC by only **1.2%** (89.0% → 87.8%) on private datasets. However, deleting M_3 makes AUROC drop sharply, because M_3 contains more dense abnormal information as a deep feature.

J. Limitations

The proposed Skip-TS model still has some limitations that will be addressed in our future work.

- 1) Although AUROC has been improved by Skip-TS in AD, it is still difficult for Skip-TS to realize accurate anomaly localization on the datasets with very small samples like BrainMRI and HeadCT. The segmentation results for these datasets, as seen in Fig. 9(g), show that the abnormal part can be located, but not fully covered, which is due to the limited training data. We aim to enhance the abnormal localization performance with very small samples in the future.
- 2) Skip-TS currently cannot detect anomalies in 3-D medical images because it relies on 2-D pre-trained networks. However, 3-D images, such as 3-D MRI [77] and 3-D CT images [78], are common in clinical practice. We will improve our method for 3-D medical images in the future.

3) The proposed AD model is exclusively trained on normal samples. When noisy anomalous samples are included in the training dataset, their performance will decline. Some supervised learning models attempt to address the problem of noisy labels. For example, in MST-TS [79], a meta-self-training method is introduced, which employs a self-training mechanism to train a T-Net and leverages the pseudo-labels generated by the teacher to train an S-Net. In the future, we will improve the robustness of our unsupervised AD model to noisy data. In particular, when the training dataset contains both normal and a certain number of anomalous samples, it is crucial to prevent the model from learning the distribution of the anomalous samples, which could lead to the incorrect detection of these normal samples as anomalies. A possible solution is to split the dataset and train several S-Nets on these sub-datasets to mitigate the influence of the noisy samples. Furthermore, incorporating the interpolated Gaussian descriptor (IGD) proposed by Chen et al. [80] into KD-based methods can help remove anomalous samples during the training phase. Additionally, training generative models like DDPM [81] to generate normal image samples for data augmentation is another potential strategy to reduce the impact of noisy samples.

V. CONCLUSION

In this article, we propose a novel AD method for medical images based on KD. To address the problem of insensitivity to anomalies caused by the structural similarity between the teacher and S-Nets, we propose an encoder-decoder architecture trained via DRKD, where the S-Net serves as a decoder and receives input directly from the T-Net. Moreover, we introduce skip connections between T-Net and S-Net to prevent the S-Net from losing normal image information across multiple scales, thereby more effectively recovering anomaly-free image representations. To further enhance the performance of our T-S model, we also design a MAC loss. Experimental results demonstrate that our method surpasses state-of-the-art by 6.4% on public datasets and 8.2% on private datasets. Additionally, compared to reconstruction-based methods, our approach achieves anomaly localization similar to that of semantic segmentation models, thus expanding the potential applications of medical AD. Future work will focus on improving the model's ability to locate anomalies and exploring additional applications in medicine.

REFERENCES

- [1] S. Pei, C. Wang, S. Cao, and Z. Lv, "Data augmentation for fMRI-based functional connectivity and its application to cross-site ADHD classification," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.
- [2] G. Yue, P. Wei, Y. Liu, Y. Luo, J. Du, and T. Wang, "Automated endoscopic image classification via deep neural network with class imbalance loss," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.
- [3] Y. Huang, G. Liu, Y. Luo, and G. Yang, "ADFA: Attention-augmented differentiable top-K feature adaptation for unsupervised medical anomaly detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2023, pp. 206–210.
- [4] H. Zhao et al., "Anomaly detection for medical images using self-supervised and translation-consistent features," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3641–3651, Dec. 2021.
- [5] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. Asian Conf. Comput. Vis.* Perth, WA, Australia: Springer, Dec. 2018, pp. 622–637.
- [6] S. Akçay, A. Atapour-Abarghouei, and T. P. Breckon, "Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2019, pp. 1–8.
- [7] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, May 2017, pp. 146–157.
- [8] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.
- [9] Y. Chen, H. Zhang, Y. Wang, Y. Yang, X. Zhou, and Q. M. J. Wu, "MAMA Net: Multi-scale attention memory autoencoder network for anomaly detection," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 1032–1041, Mar. 2021.
- [10] K. Zhou et al., "Proxy-bridged image reconstruction network for anomaly detection in medical images," *IEEE Trans. Med. Imag.*, vol. 41, no. 3, pp. 582–594, Mar. 2022.
- [11] M. Salehi, A. Eftekhari, N. Sadjadi, M. H. Rohban, and H. R. Rabiee, "Puzzle-AE: Novelty detection in images through solving puzzles," 2020, *arXiv:2008.12959*.
- [12] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, and D. V. Dylov, "Anomaly detection in medical imaging with deep perceptual autoencoders," *IEEE Access*, vol. 9, pp. 118571–118583, 2021, doi: [10.1109/ACCESS.2021.3107163](https://doi.org/10.1109/ACCESS.2021.3107163).
- [13] J. Luo, J. Lin, Z. Yang, and H. Liu, "SMD anomaly detection: A self-supervised texture-structure anomaly detection framework," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [14] K. Song, H. Yang, and Z. Yin, "Anomaly composition and decomposition network for accurate visual inspection of texture defects," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022, doi: [10.1109/TIM.2022.3196133](https://doi.org/10.1109/TIM.2022.3196133).
- [15] X. Li, J. Jing, J. Bao, P. Lu, Y. Xie, and Y. An, "OTB-AAE: Semi-supervised anomaly detection on industrial images based on adversarial autoencoder with output-turn-back structure," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [16] K. Xiao, J. Cao, Z. Zeng, and W.-K. Ling, "Graph-based active learning with uncertainty and representativeness for industrial anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023, doi: [10.1109/TIM.2023.3279422](https://doi.org/10.1109/TIM.2023.3279422).
- [17] H. Liang, L. Song, J. Du, X. Li, and L. Guo, "Consistent anomaly detection and localization of multivariate time series via cross-correlation graph-based encoder-decoder GAN," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.
- [18] R. Liu, W. Liu, H. Li, H. Wang, Q. Geng, and Y. Dai, "Metro anomaly detection based on light strip inductive key frame extraction and MAGAN network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.
- [19] M. Niu, Y. Wang, K. Song, Q. Wang, Y. Zhao, and Y. Yan, "An adaptive pyramid graph and variation residual-based anomaly detection network for rail surface defects," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
- [20] Y. Yan, D. Wang, G. Zhou, and Q. Chen, "Unsupervised anomaly segmentation via multilevel image reconstruction and adaptive attention-level transition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
- [21] C. Zhang, Y. Wang, and W. Tan, "MTHM: Self-supervised multitask anomaly detection with hard example mining," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023, doi: [10.1109/TIM.2023.3276529](https://doi.org/10.1109/TIM.2023.3276529).
- [22] H. Yang, Z. Zhu, C. Lin, W. Hui, S. Wang, and Y. Zhao, "Self-supervised surface defect localization via joint de-anomaly reconstruction and saliency-guided segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.
- [23] M. E. Tschuchnig et al., "Quantification of anomalies in rats' spinal cords using autoencoders," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104939.
- [24] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9664–9674.

- [25] J. Song, K. Kong, Y.-I. Park, S.-G. Kim, and S.-J. Kang, "Anomaly segmentation network using self-supervised learning," in *Proc. AAAI Workshop AI Design Manuf. (ADAM)*, 2022, pp. 1–16.
- [26] J. Sato et al., "Anatomy-aware self-supervised learning for anomaly detection in chest radiographs," *IScience*, vol. 26, no. 7, Jul. 2023, Art. no. 107086.
- [27] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107706.
- [28] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student–teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4183–4192.
- [29] G. Wang, S. Han, E. Ding, and D. Huang, "Student–teacher feature pyramid matching for anomaly detection," 2021, *arXiv:2103.04257*.
- [30] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14902–14912.
- [31] Y. Cao, Q. Wan, W. Shen, and L. Gao, "Informative knowledge distillation for image anomaly segmentation," *Knowl.-Based Syst.*, vol. 248, Jul. 2022, Art. no. 108846.
- [32] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9737–9746.
- [33] R. Xu et al., "Pixel-level and affinity-level knowledge distillation for unsupervised segmentation of COVID-19 lesions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1376–1380.
- [34] Q. Wan, L. Gao, and X. Li, "Logit inducing with abnormality capturing for semi-supervised image anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [35] T. Reiss, N. Cohen, L. Bergman, and Y. Hoshen, "PANDA: Adapting pretrained features for anomaly detection and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2806–2814.
- [36] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A patch distribution modeling framework for anomaly detection and localization," in *Proc. Int. Conf. Pattern Recognit.* Cham, Switzerland: Springer, Jan. 2021, pp. 475–489.
- [37] S. Lee, S. Lee, and B. C. Song, "CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization," *IEEE Access*, vol. 10, pp. 78446–78454, 2022.
- [38] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.
- [39] S. Yamada, S. Kamiya, and K. Hotta, "Reconstructed student–teacher and discriminative networks for anomaly detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 2725–2732.
- [40] M. Togaçar, Z. Cömert, B. Ergen, and Ü. Budak, "Brain hemorrhage detection based on heat maps, autoencoder and CNN architecture," in *Proc. 1st Int. Informat. Softw. Eng. Conf. (UBMYK)*, Nov. 2019, pp. 1–5.
- [41] S. Rai, S. Chowdhury, S. Sarkar, K. Chowdhury, and K. P. Singh, "A hybrid approach to brain tumor detection from mri images using computer vision," *J. Innov. Comput. Sci. Eng.*, vol. 8, no. 2, pp. 8–12, 2019.
- [42] D. Gaurav. *Covid-19 Detection*. Accessed: Jan. 10, 2023. [Online]. Available: <https://www.kaggle.com/datasets/gauravduttaikit/covid19-detection>
- [43] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863.
- [44] K. Uzair. *Breast Cancer Patients Mri's*. Accessed: Jan. 10, 2023. [Online]. Available: <https://www.kaggle.com/datasets/uzairkhan45/breast-cancer-patients-mris>
- [45] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018.
- [46] W. Wang and T. Jing. *Cp-Child.zip. Figshare. Dataset*. Accessed: Jan. 12, 2023. [Online]. Available: https://figshare.com/articles/dataset/CP-CHILD_zip/12554042/1
- [47] I. Zingman, B. Stierstorfer, C. Lempp, and F. Heinemann, "Learning image representations for anomaly detection: Application to discovery of histological alterations in drug development," 2022, *arXiv:2210.07675*.
- [48] H. Alyasriy. *The IQ-Othnced Lung Cancer Dataset*. Accessed: Jan. 14, 2023. [Online]. Available: <https://data.mendeley.com/datasets/bhmdr45bh2/1>
- [49] J. Paul Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," 2020, *arXiv:2003.11597*.
- [50] Arunava. *Malaria Cell Images Dataset*. Accessed: Jan. 12, 2023. [Online]. Available: <https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria>
- [51] M. Liu, Y. Jiao, and H. Chen, "Skip-ST: Anomaly detection for medical images using student–teacher network with skip connections," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2023, pp. 1–5.
- [52] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
- [53] I. J. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [55] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," 2020, *arXiv:2005.02357*.
- [56] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 14318–14328.
- [57] M. Xu, X. Zhou, X. Gao, W. He, and S. Niu, "Discriminative feature learning framework with gradient preference for anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, doi: [10.1109/TIM.2022.3228007](https://doi.org/10.1109/TIM.2022.3228007).
- [58] Y. Jiang, Y. Cao, and W. Shen, "A masked reverse knowledge distillation method incorporating global and local information for image anomaly detection," *Knowl.-Based Syst.*, vol. 280, Nov. 2023, Art. no. 110982.
- [59] Q. Wan, Y. Cao, L. Gao, X. Li, and Y. Gao, "Deep feature contrasting for industrial image anomaly segmentation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.
- [60] Y. Cao, Y. Zhang, and W. Shen, "High-resolution image anomaly detection via spatiotemporal consistency incorporated knowledge distillation," in *Proc. IEEE 19th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2023, pp. 1–6.
- [61] Y. Cao et al., "Segment any anomaly without training via hybrid prompt regularization," 2023, *arXiv:2305.10724*.
- [62] K. Nakazawa, K. Hotta, J. Yu, and C. Zhang, "Student–teacher anomaly detection considering knowledge consistency between layer groups," in *Proc. IEEE 11th Global Conf. Consum. Electron. (GCCE)*, Oct. 2022, pp. 381–382.
- [63] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1038–1059, Jan. 2021.
- [64] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [65] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*.
- [66] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.
- [67] H. Li and Y. Li, "Anomaly detection methods based on GAN: A survey," *Int. J. Speech Technol.*, vol. 53, no. 7, pp. 8209–8231, Apr. 2023.
- [68] P. Liznerski, L. Ruff, R. A. Vandermeulen, B. J. Franks, M. Kloft, and K. R. Müller, "Explainable deep one-class classification," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–12.
- [69] J.-H. Kim, D.-H. Kim, S. Yi, and T. Lee, "Semi-orthogonal embedding for efficient unsupervised anomaly segmentation," 2021, *arXiv:2105.14737*.
- [70] J. Yang, Y. Shi, and Z. Qi, "Learning deep feature correspondence for unsupervised anomaly detection and segmentation," *Pattern Recognit.*, vol. 132, Dec. 2022, Art. no. 108874.
- [71] G. Pang, C. Ding, C. Shen, and A. van den Hengel, "Explainable deep few-shot anomaly detection with deviation networks," 2021, *arXiv:2108.00462*.
- [72] C. Ding, G. Pang, and C. Shen, "Catching both gray and black swans: Open-set supervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7388–7398.
- [73] Y. Cai, H. Chen, X. Yang, Y. Zhou, and K.-T. Cheng, "Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images," 2022, *arXiv:2210.04227*.
- [74] M. E. Tschuchnig and M. Gadermayer, "Anomaly detection in medical imaging—A mini review," in *Proc. 4th Int. Data Sci. Conf.*, 2022, pp. 33–38.

- [75] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [76] M. Z. Khan, M. K. Gajendran, Y. Lee, and M. A. Khan, "Deep neural architectures for medical image semantic segmentation: Review," *IEEE Access*, vol. 9, pp. 83002–83024, 2021.
- [77] Q. Tian et al., "DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning," *NeuroImage*, vol. 219, Oct. 2020, Art. no. 117017.
- [78] H. Chen, X. He, H. Yang, J. Feng, and Q. Teng, "A two-stage deep generative adversarial quality enhancement network for real-world 3D CT images," *Expert Syst. Appl.*, vol. 193, May 2022, Art. no. 116440.
- [79] X. Pu and C. Li, "Meta-self-training based on teacher-student network for industrial label-noise fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.
- [80] Y. Chen, Y. Tian, G. Pang, and G. Carneiro, "Deep one-class classification via interpolated Gaussian descriptor," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 1, pp. 383–392.
- [81] A. Kazerouni et al., "Diffusion models in medical imaging: A comprehensive survey," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102846.



Mingxuan Liu (Student Member, IEEE) was born in Hunan, China, in 2002. He is currently pursuing the bachelor's degree with the School of Biomedical Engineering, Tsinghua University (THU), Beijing, China.

His research interests include small sample learning (SSL), anomaly detection, medical image analysis, generative models and spiking neural networks (SNNs).



Yunrui Jiao was born in Henan, China, in 2001. He is currently pursuing the bachelor's degree with the School of Integrated Circuits, Tsinghua University (THU), Beijing, China.

His research interests include anomaly detection, the performance of resistive random access memory (RRAM) devices and storage-in-computing.



Jingqiao Lu was born in Jingjiang, Jiangsu, China, in 1987. He received the master's degree from the School of Computer Science, Guangdong University of Technology, Guangzhou, China, in 2016.

His research interests include complex networks, big data analysis, image analysis, natural language processing, reinforcement learning, and intelligent algorithms.



Hong Chen (Senior Member, IEEE) received the Ph.D. degree from the Department of Electronic Engineering from Tsinghua University, Beijing, China, in 2005.

From 2005 to 2007, she was a Post-Doctoral Fellow at the Institute of Microelectronics of Tsinghua University (IMETU), Beijing. Since 2007, she has been with the School of Integrated Circuits, Tsinghua University, where is currently a Full Professor. From February to June in 2006, she was at the Medical Center, Nebraska University, Omaha, USA.

She was a Visiting Scholar at Gatech Tech, Atlanta, USA, from March 2016 to April 2017. She has published more than 120 journal articles and conference papers, and holds one U.S. patent and 32 Chinese patents. Her research interests include bio-inspired on-chip learning asynchronous neuromorphic circuits, and biomedical prosthetic circuits and systems.

Dr. Chen has served as an Associate Editor for IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS, a Guest Editor of Tsinghua Science and Technology, a Technical Program Committees Member of the IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC) and IEEE International Symposium on Circuits and Systems (ISCAS), and the General Chair of IEEE ASYNC 2023.