11-2021

Self-supervised multi-class pre-training for unsupervised anomaly detection and segmentation in medical images

Yu TIAN

Fengbei LIU

Guansong PANG
Singapore Management University, gspang@smu.edu.sg

Yuanhong CHEN

Yuyuan LIU

See next page for additional authors

Self-supervised Multi-class Pre-training for Unsupervised Anomaly Detection and Segmentation in Medical Images

Yu Tian[1,3*] Fengbei Liu[1*] Guansong Pang[1] Yuanhong Chen[1]
Yuyuan Liu[1] Johan W Verjans[1,2,3] Rajvinder Singh[2] Gustavo Carneiro[1]
[1] Australian Institute for Machine Learning, University of Adelaide
[2] Faculty of Health and Medical Sciences, University of Adelaide
[3] South Australian Health and Medical Research Institute

Abstract

Unsupervised anomaly detection (UAD) that requires only normal (healthy) training images is an important tool for enabling the development of medical image analysis (MIA) applications, such as disease screening, since it is often difficult to collect and annotate abnormal (or disease) images in MIA. However, heavily relying on the normal images may cause the model training to overfit the normal class. Self-supervised pre-training is an effective solution to this problem. Unfortunately, current self-supervision methods adapted from computer vision are sub-optimal for MIA applications because they do not explore MIA domain knowledge for designing the pretext tasks or the training process. In this paper, we propose a new self-supervised pre-training method for UAD designed for MIA applications, named Multi-class Strong Augmentation via Contrastive Learning (MSACL). MSACL is based on a novel optimisation to contrast normal and multiple classes of synthetised abnormal images, with each class enforced to form a tight and dense cluster in terms of Euclidean distance and cosine similarity, where abnormal images are formed by simulating a varying number of lesions of different sizes and appearance in the normal images. In the experiments, we show that our MSACL pre-training improves the accuracy of SOTA UAD methods on many MIA benchmarks using colonoscopy, fundus screening and Covid-19 Chest X-ray datasets.

1. Introduction and Background

Detecting and segmenting malignant lesions from disease screening datasets is a crucial task in medical images analysis (MIA) [2, 14, 21, 23–25, 27, 40, 42]. A challenging aspect of this problem is that such screening datasets [32, 43] contain a disproportionately large number of normal (or healthy) images, and a small minority of abnormal (or disease) images. Not only is the collection and annotation of such heavily imbalanced training sets challenging, but it is hard to acquire a representative dataset containing a reasonable number of images from all possible disease sub-classes. Instead of relying on a fully supervised approach requiring such labelled training set, we consider in this paper an alternative approach based on unsupervised anomaly detection (UAD) [7, 8, 44], which is trained exclusively with normal samples. There are two advantages with the UAD strategy: 1) the acquisition of such training set is straightforward given the large proportion of normal images in screening datasets; and 2) it is not necessary to collect a representative training set containing images from all pos-

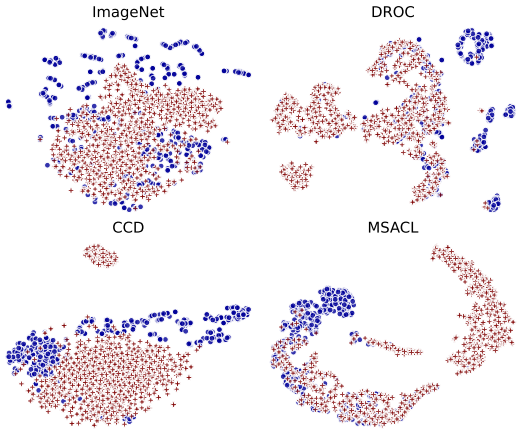

Figure 1. t-SNE results of the representations on the test set of Hyper-Kvasir [5] learned by IGD [8] after being pre-trained on ImageNet [12], or self-supervised with DROC [38], CCD [44], and our MSACL. Compared to other methods, MSACL clusters the normal image representations (blue points) in a tighter and denser region, and separates anomalous representations into three clusters (red points), which can be associated with the three classes of synthetised abnormal images formed by simulating a varying number of lesions of different sizes and appearance in the normal images.

*First two authors contributed equally to this work.

sible disease sub-classes.

UAD methods are generally based on a one-class classifier (OCC) that learns a normal image distribution from the normal training images, and test image anomalies (or abnormal images) are detected based on the extent that they deviate from the learned distribution [7, 8, 16, 26, 30, 34, 36, 37, 41, 42, 45]. UAD methods can overfit the normal class due to their strong dependence on the normal images, and a mitigating solution is based on the use of pre-training to learn representations that can be transferred to UAD. Pre-training methods can rely on ImageNet [44], which may not allow an effective representation transferring from natural images to medical images, or on self-supervised learning (SSL) [3, 6, 15, 18, 19, 22, 44], which depends on the effectiveness of the pretext tasks and the assumptions of the training process to form image classes. Self-supervision pre-training for UAD methods applied to MIA screening problems have shown promising results [44], but they have been sub-optimally explored given that they were adapted from computer vision methods without using MIA domain knowledge for designing the pretext tasks or the training process. For instance, in MIA, diseases can be divided into classes characterised by variations in the number and appearance of lesions. Moreover, previous SSL methods in UAD [38,39,44] extend contrastive learning [6] to learn representations to be transferred to multi-class classification, where representations are learned by contrasting a large number of normal images and their augmented (from geometric or appearance transformations) versions that form a large number of classes in the representation space. This is sub-optimal for MIA OCC that needs to discriminate a tight and dense cluster of normal images against a relatively small number of abnormal classes that lie outside the normal cluster (see Fig. 1).

In this paper, we propose the Multi-class Strong Augmentation via Contrastive Learning (MSACL), a new self-supervised pre-training method modelled exclusively with normal training images, and designed to learn effective image representations for different types of downstream UAD methods applied to several MIA problems. The main advantage of MSACL, compared to previous self-supervised pre-training method for MIA applications [44], is that we rely on MIA domain knowledge to design the optimisation and the pretext tasks. In particular, our optimisation uses contrastive learning to classify training samples into multiple tight and dense clusters in terms of Euclidean distance and cosine similarity, with one cluster to represent the normal images and the remaining ones to represent subclasses of the disease images (see Fig. 1). These disease images are obtained with our MedMix augmentations that simulate a varying number of lesions of different sizes and appearance in the normal training images (see Fig. 3). We summarise our contributions as follows:

- Our MSACL is the first self-supervised pre-training method specifically designed for MIA UAD applications, where our main advantage lies in the design of a contrastive learning optimisation that learns multiple classes, one for normal images, and the others for subclasses of disease images, which are formed by our MedMix augmentations that simulate a varying number of lesions of different sizes and appearance, and

- The proposed MSACL is shown to learn effective image representations that can adapt well to different types of downstream UAD methods applied to several MIA problems.

We empirically show that MSACL pre-training significantly improves the performance of two SOTA anomaly detectors, PaDiM [11] and IGD [8]. Extensive experimental results on four different disease screening medical imaging benchmarks, namely, colonoscopy images from two datasets [5, 26], fundus images for glaucoma detection [20] and Covid-19 Chest X-ray (CXR) dataset [46] show that MSACL can be used to pre-train diverse SOTA UAD methods to improve their accuracy in detecting and segmenting lesions in diverse medical images.

2. Related Work

In this section, we first discuss UAD methods, then we present self-supervised pre-training methods for computer vision and MIA problems.

UAD approaches [7, 8, 16, 26, 30, 36, 37, 41, 42, 45] can be divided into two categories: predictive-based (e.g., DSVDD [35], OC-SVM [9], and deviation network [30]), and generative-based (e.g., auto-encoder [7, 8, 16, 45] and GAN [1, 25, 36]). Predictive-based UAD approaches train a one-class classifier to describe the distribution of normal data, and discriminate abnormal data using their distance/deviation to the normal data distribution; whereas generative-based UAD approaches train deep generative models to learn latent representations of normal images, and detect anomalies based on image reconstruction error [29]. A fundamental challenge in both types of UAD methods is the learning of expressive feature representations from images, which is particularly important in MIA because abnormal medical images may have subtly looking lesions that can be hard to differentiate from normal images. Hence, if not well trained, these UAD models can overfit the normal training data and learn ineffective image representations that will fail to enable the detection and segmentation of lesions.

The representation challenge above has been addressed with the use of ImageNet [12] pre-trained models, but transferring representations learned from natural images to medical images is not straightforward [44]. Alternatively,

the representation challenge can also be tackled by self-supervised pre-training methods that learn auxiliary pretext tasks [3, 6, 15, 18, 19, 22], which is a strategy that has produced effective representations for UAD in general computer vision tasks [3, 15, 19, 39]. However, their application to MIA problems needs to be further investigated because it is not clear how to design effective optimisation functions or pretext tasks that can work well in the detection of subtle lesions in medical images. Previous UAD methods relied on self-supervised pretext tasks based on the prediction of geometric transformations [3, 15, 19] or contrastive learning using standard data augmentation techniques (e.g., scaling, cropping, etc.) [6, 18] to form a large number of image classes characterising similar and dissimilar pairs. These pretext tasks and optimisation strategy are not specifically related to the detection of subtle anomalies in medical images that contain a normal image class and a small number of disease sub-classes, so they may even degrade the detection accuracy of downstream UAD methods [47].

For SSL UAD pre-training in MIA, the only previous work CCD [44] adapts standard contrastive learning and two general computer vision pretext tasks to image anomaly detection and can be applied to multiple downstream UAD methods. Although achieving good results in many benchmarks, the optimisation explored by CCD does not explore the fact that the downstream UAD methods need to recognise one class of normal images and a small number of subclasses of disease images, and their data augmentation will not produce realistic synthetised medical image anomalies – both issues can challenge the training of downstream UAD approaches.

3. Method

In this section, we introduce the proposed MSACL pre-training approach depicted in Fig. 2. Given a training medical image dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{|\mathcal{D}|}$, with all images assumed to be from the normal class and $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^{H \times W \times C}$ (H: height, W: width, C: number of colour channels), our learning strategy involves two stages: 1) the self-supervised pre-training to learn an encoding network $f_\theta : \mathcal{X} \rightarrow \mathcal{Z}$ (with $\mathcal{Z} \subset \mathbb{R}^Z$), and 2) the fine-tuning of an anomaly detector or segmentation model built from the pre-trained $f_\theta(.)$. The approach is evaluated on a testing set $\mathcal{T} = \{(\mathbf{x}, y, \mathbf{m})_i\}_{i=1}^{|\mathcal{T}|}$, where $y \in \mathcal{Y} = \{\text{normal}, \text{abnormal}\}$, and $\mathbf{m} \in \mathcal{M} \subset \{0, 1\}^{H \times W \times 1}$ denotes the segmentation mask of the lesion in the image $\mathbf{x}$. Below, we first describe the optimisation proposed for MSACL in Sec. 3.1, then we describe the MedMix data augmentation in Sec. 3.2, followed by a bried description of the UAD methods in Sec. 3.3.

3.1. MSACL Pre-training

The gist of our proposed MSACL lies in the idea of discriminating the distribution of weakly augmented samples

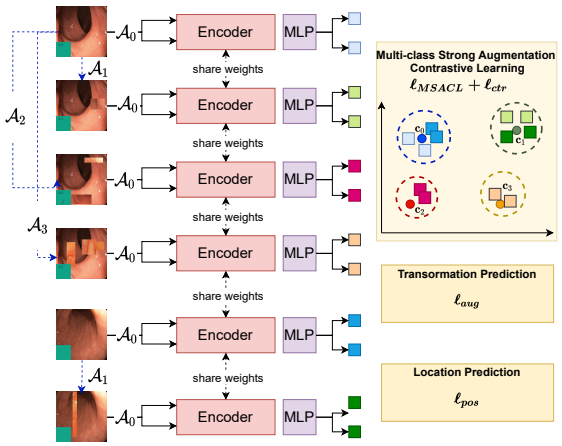

Figure 2. **MSACL**: our proposed self-supervised pre-training for UAD trains four classes of images: the normal images formed by the weak augmentations in distribution $\mathcal{A}_0$ (blue markers) and three classes of synthetised abnormal images formed by the strong augmentation in distributions $\{\mathcal{A}_n\}_{n=1}^3$ (green, pink and orange markers). The optimisation uses a constrained contrastive learning that trains a four-class classification problem. The different types of strong augmentations are produced by MedMix that introduces a varying number of fake lesions by cutting patches from the normal training images, altering them with random color jittering, Gaussian noise and non-linear intensity transformations, and pasting them to other normal training images.

(simulating normal images) from the distributions of different types of strongly augmented samples (simulating multiple classes of abnormal images). Instead of attracting and repelling samples within and between a large number of image classes [38, 39, 44], we propose a new contrastive loss to separate samples from the normal class and samples from abnormal sub-classes, and to enforce the **clusters representing the normal and abnormal sub-classes** to be **dense and tight**. To this end, our proposed loss is defined as:

$$\begin{aligned}\ell(\mathcal{D}; \theta, \beta, \gamma) =& \ell_{ctr}(\mathcal{D}; \theta) + \ell_{MSACL}(\mathcal{D}; \theta) + \\ & \ell_{aug}(\mathcal{D}; \beta) + \ell_{pos}(\mathcal{D}; \gamma),\end{aligned} \tag{1}$$

where $\ell_{ctr}(.)$ denotes the new distribution multi-centring loss, $\ell_{MSACL}(.)$ represents the new MSACL contrastive loss, $\ell_{aug}(.)$ and $\ell_{pos}(.)$ are the pretext learning losses to constrain optimisation [44], and θ, β and γ are trainable parameters. The loss terms in (1) rely on **weak data augmentation distribution**, denoted by $\mathcal{A}_0$, and $|\mathcal{A}|$ **strong data augmentation distributions**, represented by $\{\mathcal{A}_n\}_{n=1}^{|\mathcal{A}|}$, each denoting a different type of augmentation. From each of these distributions, we can sample augmentation functions $a : \mathcal{X} \rightarrow \mathcal{X}$.

The multi-centring loss in (1) depends on the estimation of the mean representation for each augmentation distribu-

tion, computed as

$$\mathbf{c}_n = \mathbb{E}_{\mathbf{x}\in\mathcal{D}, a\sim\mathcal{A}_n}[f_\theta(a(\mathbf{x}))], \tag{2}$$

where $n \in \{0, ..., |\mathcal{A}|\}$, with $\mathbf{c}_n$ being the mean representation of the training data augmented by the functions sampled from $\mathcal{A}_n$, for $n \in \{0, ..., |\mathcal{A}|\}$. Note that these mean representations are computed at the beginning of the training and frozen for the rest of the training. The distribution multi-centring loss is then defined as:

$$\ell_{ctr}(\mathcal{D}; \theta) = \mathbb{E}_{\mathbf{x}\in\mathcal{D}, n\in\{0,...,|\mathcal{A}|\}, a\sim\mathcal{A}_n} \|f_\theta(a(\mathbf{x})) - \mathbf{c}_n\|^2, \tag{3}$$

which pulls the representations of augmented samples toward their mean representations in (2), making the augmentation clusters dense and tight in Euclidean space.

To further enforce the separation between different clusters and the tightness within each cluster, we introduce a novel contrastive learning. In our contrastive learning, we maximisise the cosine similarity of samples that belong to the same class (i.e., normal or abnormal sub-classes) and minimise the cosine similarity of samples belonging to different classes. An interesting aspect of this optimisation is that samples are centred by their own cluster mean representation $\mathbf{c}_n$ from (2), so our contrastive learning, combined with the multi-centred loss in (3) will cluster samples of the same class not only in Euclidean space, but also in inner product space (with cosine measuring similarity between samples). Such re-formulated constrastive learning, combined with the multi-centring loss (3), results in a loss that produces multiple clusters, where cluster $n = 0$ contains the normal images and the others, denoted by $n \in \{1, ..., |\mathcal{A}|\}$), have the synthetised abnormal images. Our proposed MSACL loss is defined as:

$$\ell_{MSACL}(\mathcal{D}; \theta) = \mathbb{E}_{\mathbf{x}\in\mathcal{D}, n\in\{0,...,|\mathcal{A}|\}, l\in\{0,1\}} \left[\ell^x_{MSACL}(\mathbf{x}^{(n,l)}, \mathcal{D}; \theta)\right] \tag{4}$$

where $\mathbf{x}^{(n,l)} = a(\mathbf{x}^{(n)})$ represents one of two (indexed by $l \in \{0, 1\}$) augmented data obtained from the application of a weak augmentation $a \sim \mathcal{A}_0$ on a strongly augmented data denoted by $\mathbf{x}^{(n)} = a(\mathbf{x})$ with $a \sim \mathcal{A}_n$. In (4), we have:

$$\ell^x_{MSACL}(\mathbf{x}^{(n,l)}, \mathcal{D}; \theta) = -\log \frac{\exp\left[\frac{1}{\tau}\tilde{f}_\theta(\mathbf{x}^{(n,l)})^\top \tilde{f}_\theta(\mathbf{x}^{(n,(l+1) \bmod 2)})\right]}{\sum_{\substack{\mathbf{x}_j\in\mathcal{D} \\ m\in\{0,...,|\mathcal{A}|\} \\ k\in\{0,1\}}} \mathbb{I}(\mathbf{x}_j^{(m,k)} \neq \mathbf{x}^{(n,l)}) \exp\left[\kappa(n,m)\tilde{f}_\theta(\mathbf{x}^{(n,l)})^\top \tilde{f}_\theta(\mathbf{x}_j^{(m,k)})\right]}, \tag{5}$$

where $\mathbb{I}(.)$ denotes an indicator function, $\mathbf{x}_j^{(m,k)}$ is defined similarly as $\mathbf{x}^{(n,l)}$ in (4), $m \in \{0, ..., |\mathcal{A}|\}$ indexes the set of strong augmentations, and $k \in \{0, 1\}$ indexes one of the two weak augmentations applied to the strongly augmented image. Lastly, to further constrain the normal and strongly augmented data representations in (4), our MSACL loss minimises the distance between samples centred by their representation means computed as:

$$\tilde{f}_\theta(\mathbf{x}^{(n,l)}) = \frac{f_\theta(\mathbf{x}^{(n,l)}) - \mathbf{c}_n}{\|f_\theta(\mathbf{x}^{(n,l)}) - \mathbf{c}_n\|_2}, \tag{6}$$

where $\mathbf{c}_n$ is defined in (2). Also in (4) to map the representations from the same distribution into a denser region of the hyper-sphere, inspired by [10], we propose a temperature calibration strategy defined as:

$$\kappa(n, m) = \begin{cases} 1/(\alpha\tau) & \text{, if } n = m \\ 1/\tau & \text{, otherwise} \end{cases}, \tag{7}$$

where α is a scaling factor that controls the shrinkage level of the temperature τ. As a result, Eq. (7) alters the temperature for the samples belongs to the same strong augmentation distributions (i.e., when $n = m$) to a smaller value α, which allows smaller amount of repelling strength compared to samples that belongs to strong augmentation distributions $n! = m$). Putting all together, the loss in (4) clusters the image representations into hyper-spheres and regions within the hyper-spheres, where each hyper-sphere and region represent a different type of augmentation.

Inspired by [39, 44], we further constrain the training in (1) with a self-supervised classification constraint $\ell_{aug}(\cdot)$ that enforces the model to classify the strong augmentation function (Fig. 2):

$$\ell_{aug}(\mathcal{D}; \beta) = -\mathbb{E}_{\mathbf{x}\in\mathcal{D}, n\in\{0,...,|\mathcal{A}|\}, a\sim\mathcal{A}_n} \left[\log \mathbf{a}_n^\top f_\beta(f_\theta(a(\mathbf{x})))\right], \tag{8}$$

where $f_\beta : \mathcal{Z} \to [0, 1]^{|\mathcal{A}|}$ is a fully-connected (FC) layer, and $\mathbf{a}_n \in \{0, 1\}^{|\mathcal{A}|}$ is a one-hot vector representing the strong augmentation distribution (i.e., $\mathbf{a}_n(j) = 1$ for $j = n$, and $\mathbf{a}_n(j) = 0$ for $j \neq n$).

The final constraint in (1) is based on the relative patch location from the centre of the training image and is adapted for local patches (see Fig. 2). This constraint is added to learn positional and texture characteristics of the image in a self-supervised manner. Inspired by [13], the positional constraint predicts the relative position of the paired image patches, with its loss defined as

$$\ell_{pos}(\mathcal{D}; \gamma) = -\mathbb{E}_{\{\mathbf{x}_{\omega_1}, \mathbf{x}_{\omega_2}\}\sim\mathbf{x}\in\mathcal{D}} \left[\log \mathbf{p}^\top f_\gamma(f_\theta(\mathbf{x}_{\omega_1}), f_\theta(\mathbf{x}_{\omega_2}))\right], \tag{9}$$

where $\mathbf{x}_{\omega_1}$ is a randomly selected fixed-size image patch from $\mathbf{x}$, $\mathbf{x}_{\omega_2}$ is another image patch from one of its eight neighbouring patches (as shown in 'patch location prediction' in Fig. 2), $\omega_1, \omega_2 \in \Omega$ represents indices to the image lattice, $f_\gamma : \mathcal{Z} \times \mathcal{Z} \to [0, 1]^8$, and $\mathbf{p} = \{0, 1\}^8$ is a one-hot encoding of the patch location. The constraints in (8) and (9) are designed to improve training regularisation.

3.2. MedMix Augmentation

Our MedMix augmentation is designed to augment medical images to simulate multiple lesions. We target a more

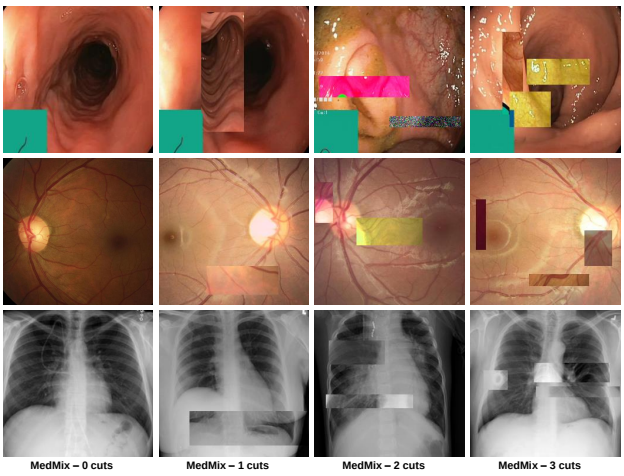

Figure 3. Examples of our MedMix data augmentation, showing augmentation $\mathcal{A}_0$ containing zero synthetic anomalies (leftmost column) and increasingly stronger augmentations $\{\mathcal{A}_n\}_{n=1}^{3}$ (second to fourth columns) with different number of synthetic anomalies (from one to three).

effective data augmentation for MIA applications than the computer vision augmentations in [44] (e.g., permutations, rotations) that do not simulate medical image anomalies and may yield poor detection performance by downstream UAD methods. We realise that anomalies in different medical domains (e.g., glaucoma and colon polyps) can be visually different, but a commonality among anomalies is that they are usually represented by an unusual growth of abnormal tissue. Hence, we propose the MedMix augmentation to simulate abnormal tissue with a strong augmentation that "constructs" abnormal lesions by cutting and pasting (from and to normal images) small and visually deformed patches. This visual deformation is achieved by applying other transformations to patches, such as colour jittering, Gaussian noise and non-linear intensity transformations. This approach is inspired by cutmix [49], where our contribution over cutmix is the intensification of the change present in the cropped patches by the appearance transformations above. These transformations are designed to encourage the model to learn abnormalities in terms of localised image appearance, structure, texture and colour.

In practice, we design $|\mathcal{A}| = 4$ strong augmentation distributions, where $\mathcal{A}_n$ includes $n \in \{0, ..., 3\}$ abnormalities in the image, which means that $\mathcal{A}_0$ denotes the normal image distribution and $\mathcal{A}_{n\in\{1,2,3\}}$ represent the abnormal image distributions, containing $\{1, 2, 3\}$ anomalous regions. Therefore, our loss targets the classification of MedMix augmentations, as shown in Fig. 3.

3.3. Anomaly Detection and Segmentation

After pre-training $f_\theta(\cdot)$ with MSACL, we fine-tune it with a SOTA UAD, such as IGD [8] or PaDiM [11]. Those methods use the same training set $\mathcal{D}$ as MSACL, containing only normal images from healthy patients.

IGD [8] combines three loss functions: 1) two reconstruction losses based on local and global multi-scale structural similarity index measure (MS-SSIM) [48] and mean absolute error (MAE) to train the encoder $f_\theta : \mathcal{X} \to \mathcal{Z}$ and decoder $g_\phi : \mathcal{Z} \to \mathcal{X}$, 2) a regularisation loss to train adversarial interpolations from the encoder [4], and 3) an anomaly classification loss to train $h_\psi : \mathcal{Z} \to [0, 1]$. The anomaly detection score of image $\mathbf{x}$ is defined by

$$s_{IGD}(\mathbf{x}) = \xi\ell_{rec}(\mathbf{x}, \tilde{\mathbf{x}}) + (1 - \xi)(1 - h_\psi(f_\theta(\mathbf{x}))), \quad (10)$$

where $\tilde{\mathbf{x}} = g_\phi(f_\theta(\mathbf{x}))$, $h_\psi(.)$ returns the likelihood that $\mathbf{x}$ is a normal image, $\xi \in [0, 1]$ is a hyper-parameter, and

$$\ell_{rec}(\mathbf{x}, \tilde{\mathbf{x}}) = \rho\|\mathbf{x} - \tilde{\mathbf{x}}\|_1 + (1 - \rho)\left(1 - (\nu m_G(\mathbf{x}, \tilde{\mathbf{x}}) + (1 - \nu)m_L(\mathbf{x}, \tilde{\mathbf{x}}))\right), \quad (11)$$

with $\rho, \nu \in [0, 1]$, $m_G(\cdot)$ and $m_L(\cdot)$ denoting the global and local MS-SSIM scores from the global and local models, respectively [8]. Anomaly segmentation uses (10) to compute $s_{IGD}(\mathbf{x}_\omega)$, $\forall\omega \in \Omega$ using global and local models, where $\mathbf{x}_\omega \in \mathbb{R}^{\hat{H}\times\hat{W}\times C}$ is an image patch. This forms a heatmap, where large values of $s_{IGD}(.)$ denote anomalous regions. The final heatmap is formed by summing up the global and local heatmaps.

PaDiM [11] utilises the multi-layer features from the pre-trained network $f_\theta(.)$ to learn a position dependent multivariate Gaussian distribution of normal image patches. Training uses samples collected from the concatenation of the multi-layer features from each patch position $\omega \in \Omega$ to learn the mean and covariance of the Gaussian model denoted by $\mathcal{N}(\mu_\omega, \Sigma_\omega)$ [11]. Anomaly detection is based on the Mahalanobis distance between the concatenated testing patch feature $\mathbf{x}_\omega$ and the learned Gaussian distribution $\mathcal{N}(\mu_\omega, \Sigma_\omega)$ at that patch position $\omega \in \Omega$ to provide a score of each patch position [11]. In particular, anomaly segmentation is inferred using the following anomaly score map:

$$s_{PaDiM}(\mathbf{x}_\omega) = \sqrt{(\mathbf{x}_\omega - \mu_\omega)^\top \Sigma_\omega^{-1}(\mathbf{x}_\omega - \mu_\omega)}, \quad (12)$$

and the final score of the whole image $\mathbf{x}$ is defined as: $s_{PaDiM}(\mathbf{x}) = \max_{\omega\in\Omega} s_{PaDiM}(\mathbf{x}_\omega)$.

4. Experiments

4.1. Datasets

We test our self-supervised pre-training MSACL on four health screening datasets, where we run experiments for

both anomaly detection and localisation. For the anomaly detection and localisation, the datasets are: the colonoscopy images of Hyper-Kvasir dataset [5], and the glaucoma dataset using fundus images [20]. For the anomaly detection, the datasets are: the colonoscopy dataset [26], and Covid-19 chest ray dataset [46] – these two datasets do not have lesion segmentation annotations, so we test anomaly detection only.

Hyper-Kvasir is a large multi-class public gastrointestinal imaging dataset [5]. We use a subset of the normal (i.e., healthy) images from the dataset for training. Specifically, 2,100 images from 'cecum', 'ileum' and 'bbps-2-3' are selected as normal, from which we use 1,600 for training and 500 for testing. We also take 1,000 abnormal images and their segmentation masks of polyps to be used exclusively for testing, with the size of 300 $\times$ 300.

LAG is a large scale fundus image dataset for glaucoma diagnosis [20]. For the experiments, we use 2,343 normal (negative glaucoma) images for training, and 800 normal images and 1,711 abnormal images with positive glaucoma with annotated attention maps by ophthalmologists in glaucoma diagnosis, with the size of 500 $\times$ 500. The attention maps is based on an alternative method for eye tracking, in which the maps are used by the ophthalmologists to explore the region of interest for glaucoma diagnosis [20].

Liu et al.'s colonoscopy dataset is a colonoscopy image dataset with 18 colonoscopy videos from 15 patients [26]. The training set contains 13,250 normal (healthy) images without polyps, and the testing set contains 967 images, with 290 abnormal images with polyps and 677 normal (healthy) images without polyps, with size of 64 $\times$ 64.

Covid-X [46] has a training set with 1,670 Covid-19 positive and 13,794 Covid-19 negative CXR images. The test set contains 400 CXR images, consisting of 200 positive and 200 negative images. We train the methods with the 13,794 Covid-19 negative CXR training images and test on the 400 CXR images, with the size of 299 $\times$ 299.

4.2. Implementation Details

For the proposed MSACL pre-training, we use Resnet18 [17] as the backbone architecture for the encoder $f_\theta(\mathbf{x})$, and similarly to previous works [6, 38], we add an MLP to this backbone as the projection head for the contrastive learning, which outputs features in $\mathcal{Z}$ of size 128. All images from the Hyper-Kvasir [5], LAG [20] and Covid-X [46] datasets are resized to 256 $\times$ 256 pixels. For the Liu et al.'s colonoscopy dataset [26], images are resized to 64 $\times$ 64 pixels. The batch size is set to 32 and learning rate to 0.01 for the self-supervised pre-training on all datasets. The model is trained using stochastic gradient descent (SGD) optimiser with momentum. For IGD and PaDiM fine-tuning, we use their default experimental hyper-parameters and settings. More details about the implementation can refer to our supplementary material.

Methods	Hyper	Liu et al.	LAG	Covid
DAE [28]	0.705	0.629	0.574	0.557
OCGAN [31]	0.813	0.592	0.534	0.612
F-anoGAN [36]	0.907	0.691	0.778	0.669
ADGAN [25]	0.913	0.730	-	-
MS-SSIM [8]	0.917	0.799	0.823	0.634
PANDA [33]	0.937	0.719	0.789	0.629
PaDiM [11]	0.923	0.741	0.688	0.614
CCD - PaDiM	0.978	0.789	0.728	0.632
MSACL - PaDiM	**0.996**	0.814	0.761	0.658
IGD [8]	0.939	0.787	0.796	0.699
CCD - IGD	0.972	0.837	0.874	0.746
MSACL - IGD	0.995	**0.851**	**0.908**	**0.872**

Table 1. **Anomaly detection: AUC** test results on Hyper-Kvasir, Liu et al.'s colonocopy, LAG and Covid-X, respectively.

4.3. Evaluation Measures

The anomaly detection performance is quantitatively assessed by the area under the ROC curve (AUROC). For anomaly segmentation, the performance is measured by Intersection over Union (IoU) and Dice score.

4.4. Lesion Detection Results

In Table 1, we show the results of anomaly detection on Hyper-Kvasir, LAG, Liu et al.'s colonoscopy and Covid-X dataset, where we present results from baseline UAD methods, including OCGAN [31], f-AnoGAN [36], ADGAN [26], and deep autoencoder [28] and its variant with MS-SSIM loss [8]. We also show the results the current SOTA UAD baseline PANDA [33]. As discussed in Sec. 3.3, we choose IGD [8] and PaDiM [11] as the anomaly detector for evaluating our proposed MSACL pre-training approach and compare it with previous SOTA SSL CCD [44] approach to fine-tune IGD and PaDiM.

Detection Results on Hyper-Kvasir. For Hyper-Kvasir, comparing with the baseline UAD methods, the performance of PaDiM and IGD are improved using our MSACL pre-trained encoder by around 5% and 6% AUC, which achieves SOTA anomaly detection AUC results of 99.6% and 99.5%, respectively, on Hyper-Kvasir. Comparing with current SOTA CCD pre-training [44], our MSACL pre-training improves the performance by 2.3% and 1.8% for PaDiM and IGD, on Hyper-Kvasir.

Detection Results on LAG. For LAG dataset, our MSACL pre-training improves PaDiM and IGD AUCs by 7.3% and 11.2%, compared with their ImageNet pre-trained model, where the MSACL pre-trained IGD achieves the SOTA results of 90.8% AUC. Comparing with CCD pre-trained PaDiM and IGD [44], our proposed MSACL pre-trained PaDiM and IGD surpass them by 3.3% and 3.4% in

terms of AUC. For LAG, the glaucoma features are often subtle and easy to neglect, so IGD with both reconstruction and anomaly classification constraints can generally perform better than PaDiM variants.

Detection Results on Liu et al.'s colonoscopy dataset. We further test our approach on Liu et al.'s colonoscopy dataset [25], as shown in Table 1. Similarly to the previous results, our MSACL pre-trained PaDiM improves the ImageNet pre-trained PaDiM by 7.3% AUC, and CCD pre-trained PaDiM by 2.5% of AUC. The IGD with the MSACL pre-trained encoder achieves the SOTA result of 85.1% AUC, surpassing the previous CCD and ImageNet pre-trained IGD by 1.4% and 6.4% AUC, respectively. The gap between PaDiM and IGD may be due to the low resolution of the images in this dataset, which hinders the PaDiM performance that requires dense intermediate feature maps.

Detection Results on Covid-X. Our MSACL pre-trained PaDiM and IGD methods achieve 65.8% and 87.2% AUC on the Covid-X dataset, significantly surpassing their ImageNet pre-trained by 4.4% and 17.2% AUC, and CCD pre-trained by 2.6% and 12.6% AUC. The small abnormal lesions in CXR images are hard to detect, so the generative-based anomaly detector IGD can learn more effectively the fine-grained appearances of normal images, leading to better generalisation to detect unseen anomalous regions during testing with the SOTA results of 87.2% AUC.

Finally, the performances of the baseline UAD methods are well below our self-supervised MSACL pre-training with IGD and PaDiM on all four datasets. All results from MSACL pre-trained PaDiM and IGD show that our proposed MedMix and MSACL loss improve the generalisation of the fine-tuning stage for anomaly detection and produce better constrained feature space of normal samples. The SOTA performance across four different medical imaging datasets indicate that our pre-training is a general pre-training approach that is adaptable to different medical domains (i.e., CXR, colonoscopy, fundus). Moreover, achieving SOTA results on two different types of anomaly detectors suggests that our self-supervised pre-training can produce good representations for both generative and predictive anomaly detectors, demonstrating the applicability of MSACL to real-world clinical systems.

4.5. Lesion Segmentation Results.

Segmentation results on Hyper-Kvasir. We demonstrate the anomaly segmentation performance on Hyper-Kvasir on Table 2. Following [44], we randomly sample 100 abnormal images from the test set and compute the mean segmentation performance over five different such groups of 100 images. The proposed MSACL pre-training improves the IGD and PaDiM by 1.2% and 2.8% IoU compared with the CCD pre-training, and 8.1% and 6.4% IoU with respect to the ImageNet pre-training, respectively. In addition, our MSACL pre-trained PaDiM shows the SOTA result of 40.6% IoU and 55.4% Dice, demonstrating the effectiveness of our MSACL approach for abnormal lesion segmentation.

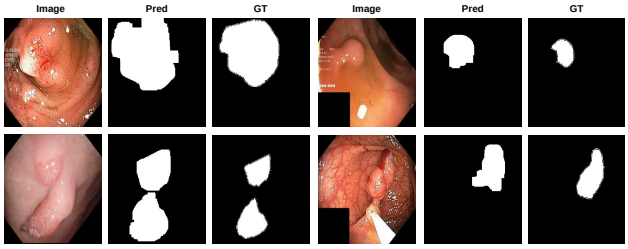

Figure 4. Segmentation of four abnormal images from Hyper Kvasir [20], with their predictions (Pred) and ground truth annotations (GT), using PaDiM with MSACL pre-training.

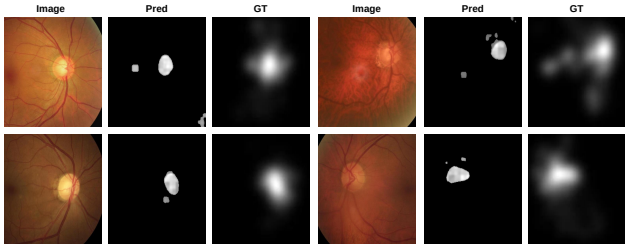

Figure 5. Segmentation of four abnormal images from LAG [20], with their predictions (Pred) and ground truth attention maps (GT), using IGD with MSACL pre-training.

Methods	Hyper		LAG	
	IoU	Dice	IoU	Dice
PaDiM [11]	0.341	0.475	0.427	0.579
CCD - PaDiM	0.378	0.497	0.462	0.612
MSACL - PaDiM	**0.406**	**0.554**	0.475	0.643
IGD [8]	0.303	0.417	0.409	0.539
CCD - IGD	0.372	0.502	0.509	0.645
MSACL - IGD	0.384	0.521	**0.516**	**0.667**

Table 2. **Anomaly segmentation:** Mean IoU, Dice results on test set of **Hyper-Kvasir** on 5 different groups of 100 images with ground truth masks, and on abnormal samples from **LAG** test set. Best results for each case are highlighted.

Segmentation results on LAG. We further demonstrate the segmentation results on LAG dataset on Table 2. The MSACL pre-trained IGD achieves the SOTA result of 51.6% IoU, 66.7% Dice, showing that our model can effectively segment different types of lesions, such as colon polyps or optic disk and cup with Glaucoma. Moreover, PaDiM pre-trained with MSACL improves PaDiM pre-trained with CCD and ImageNet by 1.3% and 4.8% IoU, respectively. Also, PaDiM with MSACL pre-training achieves 64.3% Dice, which are comparable to the SOTA results by the MSACL pre-trained IGD.

CCD [44]	MedMix	ℓ_{MSACL}	$\kappa(n,m)$	AUC - Hyper	AUC - LAG
✓				0.978	0.728
✓	✓			0.985	0.739
	✓	✓		0.993	0.753
	✓	✓	✓	0.996	0.761

Table 3. **Ablation study of the MSACL components** on test set of Hyper-Kvasir and LAG, using PaDiM [11] as anomaly detector.

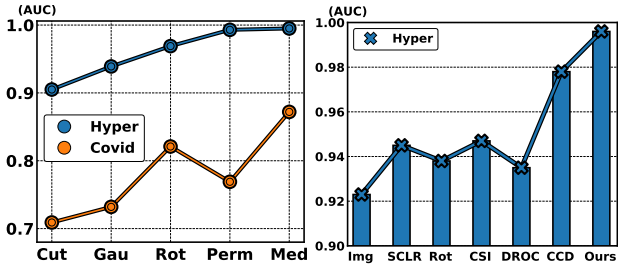

Figure 6. **Left:** Anomaly detection testing results in terms of different types of strong augmentations (i.e., Cutmix, Gaussian noise, Rotation, Permutation, and our MedMix) on Hyper-Kvasir and Covid-X, where IGD [8] is used as the anomaly detector. **Right:** Ablation studies with different self-supervised pre-training approaches (i.e., ImageNet, SimCLR [6], Rot-Net [15], CSI [39], DROC [38], CCD [44] and our MSACL) on Hyper-Kvasir testing set, where PaDiM [11] is used as the anomaly detector.

4.6. Qualitative Results

Visualisation of predicted segmentation. The visualisation of polyp segmentation results of PaDiM with MSACL pre-training on Hyper-Kvasir [5] is shown in Fig. 4. Notice that our model can effectively segment colon polyps with various sizes and shapes. We also show the segmentation results based on the pixel-level anomaly scores of IGD with MSACL pre-training on the LAG dataset in Fig. 5.

Visualisation of t-SNE results. To validate our proposed MSACL pre-training, we compare the image representations produced by ImageNet, CCD, DROC and MSACL pre-training, using t-SNE on Hyper-Kvasir. The results are shown in Fig. 1. The proposed MSACL appears to cluster all the normal data into a denser and tighter region of the representation space, where the abnormal data fall outside of this region in relatively distinct three clusters. In contrast, the models pre-trained with the other approaches produce a poorly clustered normal data that is likely to challenge the training of the downstream UAD method.

4.7. Ablation Study

Different components of MSACL pre-training. We present an ablation study that shows the influence of each step of our proposed MSACL pre-training and PaDiM fine-tuning in Table 3 on Hyper-Kvasir and LAG datasets. Starting from CCD [44], we notice that the use of MedMix can improve the AUC on both datasets by 1%. Replacing the CCD by the MSACL loss provides another improvement of between 1% and 2%. Then, adding the temperature calibration $\kappa(n,m)$ from (7) provides around 0.5% improvement.

Different strong augmentations. In Fig. 6 - left, we explore the influence of strong augmentation strategies, represented by rotation, permutation, cutout, Gaussian noise and our proposed MedMix on the AUC results on Hyper-Kvasir and Covid-X datasets, based on our self-supervised MSACL pre-training with IGD as anomaly detector. The performance of our MedMix reaches the SOTA results of 99.5% and 87.2% on those datasets. The second best AUC (96.9%) on Hyper-Kvasir uses random permutations, which were used in CCD pre-training [44], producing an AUC 0.2% worse than our MedMix. For Covid-X, rotation is the second best data augmentation approach with an AUC result that is 5.1% worse than MedMix. Other approaches do not work well with the appearance characteristics of X-ray images, yielding significantly worse results than our MedMix on Covid-X. These results suggest that the use of MedMix as the strong augmentation yields the best AUC results on different medical image benchmarks.

Different self-supervised methods. In Fig. 6 - right, we show the results on Hyper-Kvasir from different pre-training approaches, using PaDiM as anomaly detector. It can be observed that our MSACL approach surpasses previous SOTA CCD pre-training [44] by 2.2% AUC. Other pre-training methods proposed in computer vision (e.g., ImageNet pre-training, SimCLR [6], Rot-Net [15]) achieve worse results than CCD and MSACL. An interesting point in this comparison is the poor result from ImageNet pre-training, suggesting that it may not generalise well for anomaly detection in medical images. Finally, our MSACL achieves better results than previous SOTA UAD SSL approaches CSI [39] and DROC [38] by about 4% $\sim$ 5% AUC, indicting the effectiveness of our new contrastive loss.

5. Conclusions and Discussions

In this paper, we proposed a new self-supervised pre-training approach, namely MSACL, for UAD methods applied to MIA problems. MSACL is based on a new contrastive learning optimisation to learn multiple classes of normal and abnormal images, formed with the proposed MedMix data augmentation that simulates medical abnormalities. After pre-training a UAD model using our MSACL, we fine-tune it with two SOTA anomaly detecting approaches. The experimental results indicate that our MSACL pre-training can effectively improve the performance of anomaly detection and segmentation on several medical datasets for both anomaly detectors. In the future, we plan to design a new anomaly detector that suits better the characteristics of our self-supervised MSACL pre-training.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018. 2

[2] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Scale-space autoencoders for unsupervised anomaly segmentation in brain mri. In *MICCAI*, pages 552–561. Springer, 2020. 1

[3] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*, 2020. 2, 3

[4] David Berthelot, Colin Raffel, Aurko Roy, and Ian Goodfellow. Understanding and improving interpolation in autoencoders via an adversarial regularizer. *arXiv preprint arXiv:1807.07543*, 2018. 5

[5] Hanna Borgli and et al. Hyperkvasir, a comprehensive multiclass image and video dataset for gastrointestinal endoscopy. *Scientific Data*, 7(1):1–14, 2020. 1, 2, 6, 8

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020. 2, 3, 6, 8

[7] Xiaoran Chen, Suhang You, Kerem Can Tezcan, and Ender Konukoglu. Unsupervised lesion detection via image restoration with a normative prior. *Medical image analysis*, 64:101713, 2020. 1, 2

[8] Yuanhong Chen, Yu Tian, Guansong Pang, and Gustavo Carneiro. Unsupervised anomaly detection and localisation with multi-scale interpolated gaussian descriptors. *arXiv preprint arXiv:2101.10043*, 2021. 1, 2, 5, 6, 7, 8

[9] Yunqiang Chen, Xiang Sean Zhou, and Thomas S Huang. One-class svm for learning in image retrieval. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 34–37. IEEE, 2001. 2

[10] Hyunsoo Cho, Jinseok Seol, and Sang-goo Lee. Masked contrastive learning for anomaly detection. *arXiv preprint arXiv:2105.08793*, 2021. 4

[11] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. *arXiv preprint arXiv:2011.08785*, 2020. 2, 5, 6, 7, 8

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 1, 2

[13] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *ICCV*, pages 1422–1430, 2015. 4

[14] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *MICCAI*, pages 263–273. Springer, 2020. 1

[15] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. *arXiv preprint arXiv:1805.10917*, 2018. 2, 3, 8

[16] Dong Gong et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *ICCV*, pages 1705–1714, 2019. 2

[17] Kaiming He et al. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 6

[18] Kaiming He et al. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 2, 3

[19] Dan Hendrycks et al. Using self-supervised learning can improve model robustness and uncertainty. *arXiv preprint arXiv:1906.12340*, 2019. 2, 3

[20] Liu Li et al. Attention based glaucoma detection: A large-scale database and cnn model. In *CVPR*, pages 10571–10580, 2019. 2, 6, 7

[21] Geert Litjens et al. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017. 1

[22] Fengbei Liu et al. Self-supervised depth estimation to regularise semantic segmentation in knee arthroscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 594–603. Springer, 2020. 2, 3

[23] Fengbei Liu, Yu Tian, Filipe R Cordeiro, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. Noisy label learning for large-scale medical image classification. *arXiv preprint arXiv:2103.04053*, 2021. 1

[24] Fengbei Liu, Yu Tian, et al. Self-supervised mean teacher for semi-supervised chest x-ray classification. *arXiv preprint arXiv:2103.03629*, 2021. 1

[25] Y. Liu, Y. Tian, G. Maicas, L. Z. Cheng Tao Pu, R. Singh, J. W. Verjans, and G. Carneiro. Photoshopping colonoscopy video frames. In *ISBI*, pages 1–5, 2020. 1, 2, 6, 7

[26] Y. Liu, Y. Tian, G. Maicas, L. Z. Cheng Tao Pu, R. Singh, J. W. Verjans, and G. Carneiro. Photoshopping colonoscopy video frames. In *ISBI*, pages 1–5, 2020. 2, 6

[27] Cheng Tao Pu LZ et al. Computer-aided diagnosis for characterisation of colorectal lesions: a comprehensive software including serrated lesions. *Gastrointestinal Endoscopy*, 2020. 1

[28] Jonathan Masci and et al. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks*, pages 52–59. Springer, 2011. 6

[29] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021. 2

[30] Guansong Pang, Chunhua Shen, and Anton van den Hengel. Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 353–362, 2019. 2

[31] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *CVPR*, pages 2898–2906, 2019. 6

[32] L Pu, Zorron Cheng Tao, et al. Prospective study assessing a comprehensive computer-aided diagnosis for characterization of colorectal lesions: results from different centers and imaging technologies. In *Journal of Gastroenterology and Hepatology*, volume 34, pages 25–26. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019. 1

[33] Tal Reiss, Niv Cohen, Liron Bergman, and Yedid Hoshen. Panda: Adapting pretrained features for anomaly detection and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2806–2814, 2021. 6

[34] Tal Reiss and Yedid Hoshen. Mean-shifted contrastive loss for anomaly detection. *arXiv preprint arXiv:2106.03844*, 2021. 2

[35] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018. 2

[36] Thomas Schlegl et al. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 2, 6

[37] Philipp Seeböck, José Ignacio Orlando, Thomas Schlegl, Sebastian M Waldstein, Hrvoje Bogunović, Sophie Klimscha, Georg Langs, and Ursula Schmidt-Erfurth. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal oct. *IEEE transactions on medical imaging*, 39(1):87–98, 2019. 2

[38] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, and Tomas Pfister. Learning and evaluating representations for deep one-class classification. *arXiv preprint arXiv:2011.02578*, 2020. 1, 2, 3, 6, 8

[39] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *arXiv preprint arXiv:2007.08176*, 2020. 2, 3, 4, 8

[40] Yu Tian et al. One-stage five-class polyp detection and classification. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 70–73. IEEE, 2019. 1

[41] Yu Tian et al. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. *arXiv preprint arXiv:2101.10030*, 2021. 2

[42] Yu Tian, Gabriel Maicas, Leonardo Zorron Cheng Tao Pu, Rajvinder Singh, Johan W Verjans, and Gustavo Carneiro. Few-shot anomaly detection for polyp frames from colonoscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 274–284. Springer, 2020. 1, 2

[43] Yu Tian and otherss. Detecting, localising and classifying polyps from colonoscopy videos using deep learning. *arXiv preprint arXiv:2101.03285*, 2021. 1

[44] Yu Tian, Guansong Pang, Fengbei Liu, Seon Ho Shin, Johan W Verjans, Rajvinder Singh, Gustavo Carneiro, et al. Constrained contrastive distribution learning for unsupervised anomaly detection and localisation in medical images. *MICCAI 2021*, 2021. 1, 2, 3, 4, 5, 6, 7, 8

[45] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *ECCV*, pages 485–503. Springer, 2020. 2

[46] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12, 2020. 2, 6

[47] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *ICML*, pages 9929–9939. PMLR, 2020. 3

[48] Zhou Wang and et al. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 5

[49] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6023–6032, 2019. 5