Self-supervised Anomaly Detection with Random-shape Pseudo-outliers

Hanqiu Deng and Xingyu Li

Abstract— **Anomaly detection in a medical image is a challenging yet essential task. It relies on learning patterns/distributions from health data only, and no abnormal samples are available during training. This study proposes a novel self-supervised learning method to precisely detect and localize anomalies in MRI medical images. We synthesize abnormal images by overlaying random pseudo-outliers onto normal samples and propose a discriminative model for anomaly detection. Unlike prior arts that generate abnormalities with pre-determined regular geometric shapes, we introduce a new outlier synthesis strategy capable of generating random-shape anomalies. By learning the disentanglement of pseudo-outliers and normal regions in the synthesized images, our model can capture natural anomalies in images at both the pixel level and sample level. We present our empirical experimentation on two publicly accessible datasets and demonstrate the proposed method's superiority over SOTA solutions on MRIs.**

I. Introduction

Automatic detection of anomalies such as tumors in medical images is one of the essential tasks in computational medical imaging. Initially, this task was tackled by supervised learning with pairs of medical images with abnormalities and corresponding manual annotations as their ground truth. For instance, to localize tumors in brain MRI images, many efforts required the manual annotation of tumors by radiologists for model training [1]. Although supervised learning-based methods have delivered plausible results, their performance heavily relies on the quality of annotated data. In addition, the generalizability of supervised anomaly detection models is in question on unseen lesions or abnormalities. To overcome the foregoing problems, many studies propose the use of unsupervised learning techniques for anomaly detection. By learning from normal medical images from healthy patients, a model is expected to discriminate any abnormalities in query images.

Since annotation of abnormal samples is unavailable in the setting of unsupervised anomaly detection, many efforts tackle this problem through constructing self-supervision tasks over the provided normal images. Particularly, learning health features for normal image reconstruction is a plausible way for this purpose. Recently, various generative models such as Variational Auto-Encoders (VAE) [2] and Generative Adversarial Networks (GAN) [3] have demonstrated powerful capabilities in image generation and been applied to explore the healthy patterns in unsupervised anomaly detection. In theory, since a generative model is exposed to normal images only, it has limited capability to restore samples with abnormality from their low-dimensional encoding. Consequently, normal samples have smaller reconstruction residues, whereas anomalous images lead to more significant restoration errors. That is, image regions with large residues are strong candidates to anomalies [5], [6], [4]. However, the performance of such generative model-based anomaly detection methods heavily depends on the quality of the reconstructed samples and the generalization ability of the models. It is observed that many generative models preserve anomalous content in the restoration output when an abnormal region is large in an image. This unexpected failure is usually accompanied by a small restoration residue, which may lead to anomaly misidentification.

Alternatively, several studies attempt to construct self-supervision segmentation or image-filling tasks by synthesizing abnormalities in normal images. Learning the representation of unlabelled or single-class dataset by falsifying the data is also defined as self-supervised learning [16]. Inspired by the context encoders for image repair [13], Zimmerer et al. proposed to mask a square region as anomalies in a brain MRI and utilized a context-encoding VAE (ceVAE) to restore the missing content [7]. Foreign patch interpolation (FPI) [14] is another way to synthesize subtle defects for anomaly detection on Brain MRIs. In contrast to anomaly elimination with generative model, FPI is a discriminative approach to directly detect synthetic anomalies. Originating from FPI [14], Tan et al. introduced the PII strategy to modify the content of an image patch in chest X-ray and fetal ultrasound, which perturbs the spatial structures in the original normal samples. By formulating anomaly detection as a pseudo-supervised segmentation framework with this PII strategy, an Encoder-Decoder model is optimized by minimizing the pixel-level binary cross-entropy loss. We argue that in addition to the random numerical variation, the diversity of shapes of synthesized abnormal regions is also important. Regular geometric shapes of synthetic abnormalities in prior studies hinder the precision detection of natural anomalies.

In this paper, we propose a novel self-supervised strategy for anomaly detection in MRI images. To synthesize pseudo-outliers with irregular shapes, we particularly introduce an algorithm to generate random-shape masks. Three intuitive abnormality synthesis functions are introduced. During training, a normal image is perturbed by a randomly generated pseudo-outlier, and the combo loss [15] (i.e. combination of Dice Loss and Binary Cross-Entropy loss) is used to evaluate the agreement between the predicted anomaly score map and corresponding mask. Our contributions in this study are summarized as follows. First, we design a strategy for generating random-shape outliers for anomaly detection.

Hanqiu Deng and Xingyu Li are with Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. hanqiu1@ualberta.ca, xingyu@ualberta.ca

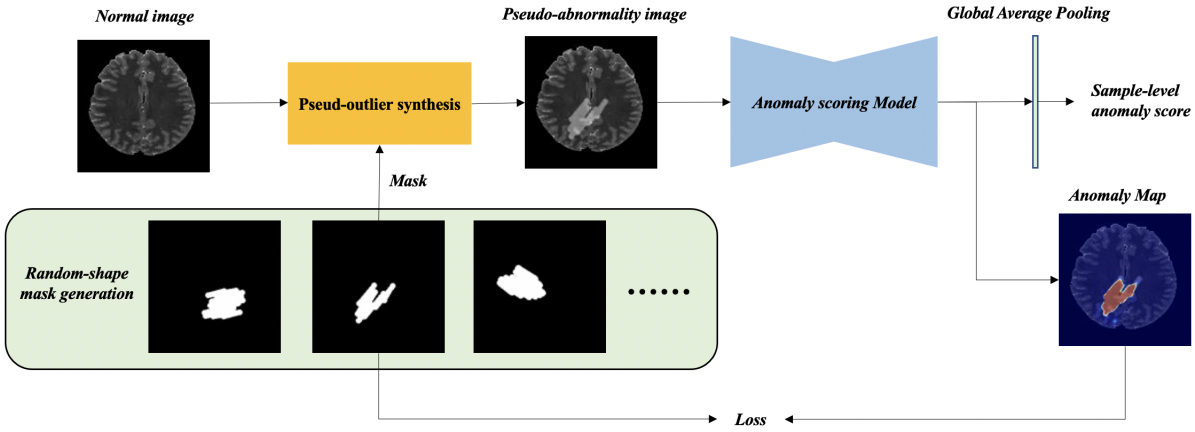

Fig. 1. Diagram of the proposed solution, where a heatmap is used to visualize the degree of abnormality in an image. This study introduces a random-shape pseudo-outlier synthesis strategy to blend artificial abnormalities into normal MRI images. We adopt the encoder-decoder architecture to estimate anomaly scores at pixel level and sample level. During inference, a query image is directly passed to the anomaly scoring model for its anomaly score.

Second, we adopt the combo loss to train an anomaly scoring model. Finally, we demonstrate the effectiveness of this approach in anomaly detection on public medical imaging datasets.

II. Methodology

Fig.1 depicts the diagram of the proposed anomaly detection solution. Specifically, in the setting of anomaly detection, there are no abnormal images available. Therefore, we introduce an outlier synthesis strategy to construct pseudo supervision data. Given a set of healthy images, $\mathcal{I} = \{\mathbf{I}_1, ..., \mathbf{I}_N\}$, we synthesize a new image set $\{\widetilde{\mathcal{I}}, \mathcal{M}\} = \{(\widetilde{\mathbf{I}}_1, \mathbf{M}_1), ..., (\widetilde{\mathbf{I}}_N, \mathbf{M}_N)\}$, where $\mathbf{M}_i$ is the binary anomaly mask used to generate anomalous synthesis $\widetilde{\mathbf{I}}_i$. In this paper, we use boldface letters to represent matrices and vectors. Taking the anomaly mask as the ground truth, we successfully convert the anomaly detection into a self-supervised problem. That is, given the synthesis training set $\{\widetilde{\mathcal{I}}, \mathcal{M}\}$, we train an anomaly scoring model $f_\theta : \widetilde{\mathcal{I}} \rightarrow \mathcal{M}$. The sample-level anomaly score is calculated by averaging the entire anomaly map. The rest of this section will elaborate on our random-shape pseudo-outlier synthesis strategy and anomaly scoring model.

A. Random-shape Pseudo-Outlier Synthesis

Synthetic anomalies in literature are usually constrained within rectangular or square patches. However, most real anomalies in medical images, for example, tumors, are in irregular shapes. Fig. 3 shows examples of irregular-shape anomalous regions in the BRATS MRI image set. To bridge this gap, we propose a random-shape pseudo-outlier synthesis strategy consisting of two modules: random-shape mask generation and abnormality blend-in. A mask $\mathbf{M}$ defines the shape, size, and location of a synthetic anomaly, and the blend-in module focuses on abnormal image generation.

Our random-shape mask generation algorithm originated from [9]. It utilizes multiple random lines to imitate a paintbrush to mask regions in images. To obtain a compact, solid mask for downstream anomaly blend-in, we improve the original algorithm by introducing more constraints on line directions and spacing. Briefly, our algorithm starts with drawing a straight line of an arbitrary length, followed by another straight line at a small angle toward the opposite direction. We repeat the above procedure several times for a random-shape mask. The specific mask generation algorithm is presented in Algorithm.1.

Algorithm 1 : Random-shape mask generation

$mask = zeros(image\_height, image\_width)$
$NumPoints = random.uniform(MinPoints, MaxPoints)$
$PointX = random.uniform(0, image\_height)$
$PointY = random.uniform(0, image\_width)$
$BrushWidth = random.uniform(MinWidth, MaxWidth)$
$angle = random.uniform(0, 2\pi)$
for $i = 0$ *to* $NumPoints$ **do**
  if $i\%2 == 0$ **then**
    $angle = angle + (1 + \sigma)\pi$
  else
    $angle = angle - (1 + \sigma)\pi$
  end if
  $length = random.uniform(MinLength, MaxLength)$
  $DrawLine(mask, (PointX, PointY), BrushWidth, angle, length)$
  $PointX = PointX + length * sin(angle$
  $PointY = PointY + length * cos(angle)$
  $DrawCircle(mask, (PointX, PointY), BrushWidth//2)$
end for
$mask = random.flipLeftRight(mask)$
$mask = random.flipLeftRight(mask)$

With the obtained anomaly mask, we proceed to synthesize abnormal MRIs. We hypothesize that all normal MRI slices follow similar healthy anatomical distributions and any disturbance to the spatial structure may result in anomalies. Therefore, we vary pixel values within a mask region and blend the synthetic anomalies into the original normal MRI image. In this study, we consider three transformations to

implement the interference:

$$\dot{\mathbf{I}} = \begin{cases} g(\mathbf{I}+\mathbf{N}), & \mathbf{N} \sim \mathcal{N}(0,1) \\ g(\mathbf{I}+c), & c \sim \mathcal{N}(0,1), \\ g(\mathbf{I}+\mathbf{I}\cdot c), & c \sim \mathcal{N}(0,1) \end{cases} \tag{1}$$

where $g(\cdot)$ represents a clip function that limits image values in the range of $[0,1]$, $\mathbf{N}$ is a Gaussian noise image, and c is constant scalar randomly drawn from the Gaussian distribution $\mathcal{N}(0,1)$. In practice, we regenerate any $\dot{\mathbf{I}}$ if $\sup|\dot{\mathbf{I}}-\mathbf{I}| < \varepsilon$. During training, we randomly select one of these transformations for each feeding sample. The synthetic abnormal image $\widetilde{\mathbf{I}}$ is then generated by blending the pseudo-outlier $\dot{\mathbf{I}}$ into its original sample $\mathbf{I}$:

$$\widetilde{\mathbf{I}} = (\mathbf{1}-\mathbf{M})\cdot\mathbf{I} + \mathbf{M}\cdot\dot{\mathbf{I}}. \tag{2}$$

Note that our aim in synthesising pseudo-outlier data is not to estimate the true distribution of lesions. In contrast to previous work [14], we focus on the utilizing the randomness of synthetic data to guide the model to identify healthy image regions. During the training process, the healthy region samples are constant, while the synthetic regions consistently vary randomly. To cope with unknown changes, the model tends to identify healthy regions different from unknown anomalies. We apply three transformations as random numerical variations, which are related to noise, luminance and contrast. We also propose the random-shape synthesis to enhance the effect of anomaly localization.

B. Anomaly Scoring Model

To estimate pixel-level anomaly scores, we propose training a U-Net structured network on the synthesis dataset to disentangle outliers and normal context. Fig.2 depicts the specific network architecture in this study. In the downsampling path, each module has two convolutional layers and a MaxPooling layer. The bottleneck block is composed of two convolutional layers. In the upsampling path, each block consists of a deconvolutional layer and two convolutional layers. In the output layer, a softmax function is used to normalize the $2 \times 128 \times 128$ logits into anomaly scores in the range of $[0,1]$.

A good anomaly scoring model should (1) never miss anomalies and (2) assign high scores to anomalous pixels. In this regard, we propose to train our model f_θ using the combo loss to measure the agreement between the estimation $f_\theta(\widetilde{\mathbf{I}})$ and the synthesis mask M:

$$\mathcal{L}_{ce}(f_\theta(\widetilde{\mathbf{I}}), \mathbf{M}) = \alpha\cdot\mathcal{L}_{dice} + (1-\alpha)\cdot\mathcal{L}_{bce}, \tag{3}$$

where α is a hyperparameter to balance the dice loss $\mathcal{L}_{dice}$ and pixel-level binary cross-entropy loss (BCE) $\mathcal{L}_{bce}$. Dice loss measures the similarity between two regions, helping the model localize abnormalities in an image precisely:

$$\mathcal{L}_{dice} = 1 - \frac{2\sum[f_\theta(\widetilde{\mathbf{I}})\cdot\mathbf{M}]}{\sum f_\theta(\widetilde{\mathbf{I}}) + \sum\mathbf{M}}. \tag{4}$$

The BCE loss is a distribution-based loss over all pixels. It measures the similarity between two distributions but without considering any spatial information in the score estimation.

$$\mathcal{L}_{bce} = -\sum[\mathbf{M}\cdot log(f_\theta(\widetilde{\mathbf{I}})) + (\mathbf{1}-\mathbf{M})\cdot log(\mathbf{1}-f_\theta(\widetilde{\mathbf{I}})]. \tag{5}$$

In the inference stage, a query image is directly fed into the network f_θ for its anomaly score.

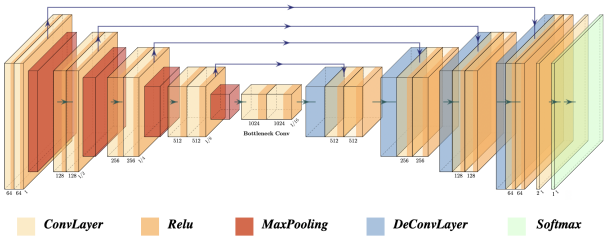

Fig. 2. Our anomaly scoring model adopts the U-Net structure. There are four downsampling layers, each consisting of two convolutional layers and a maximum pooling layer. The bottleneck layer is two convolutional layers. Correspondingly, the upsampling layer consists of two convolutional layers and a deconvolutional layer. The number of output channels we have chosen is $[32, 64, 128, 256, 512, 256, 128, 64, 32]$. The softmax layer converts pixel-level logits from a final convolutional layer into anomaly probability values in the range of $[0,1]$.

III. Experimentation and Discussions

A. Implementation and evaluation details:

In random-shape mask generation, (MinPoints, MaxPoints), (MinWidth, MaxWidth), and (MinLength, MaxLength) are set to (8, 32), (4,16), and (8, 24), respectively. Similar to conventional data augmentation, our pseudo-outlier images are generated during the training processing. We set $\alpha = 0.5$ in the loss function and take Adam as the optimizer for our anomaly scoring model, with a learning rate of 1e-4. Our model are trained 10 epoches with a batch size of 16.

We follow previous studies and use the Area Under the Receiver Operating Characteristic curve (AUROC) and Average Precision (AP) as evaluation metrics in both sample level and pixel level.

B. Experiments on BRATS Brain MRI dataset

BRATS Brain MRI dataset: We follow the convention in prior arts [4], [17], [7] and construct the anomaly detection dataset from two public Brain MRI datasets: Human Connectome Project Young Adult (HCP) [10] and Multimodal Brain Tumor Image Segmentation benchmark (BRATS) 2020 [1]. Concretely, HCP provides 45 T2-weighted brain MRI voxels collected from healthy patients. For each MRI voxel, we extract 90 complete central slices. The entire 4050 normal MRI images obtained from the HCP dataset are used for training. BRATS is a dataset providing pixel-level tumor segmentation annotation. We extract 80 slices from each T2-weighted Brain MRI voxel and its corresponding segmentation ground truth. The 368 T2-weighted Brain MRI voxels in BRATS result in 29440 images for testing. All images are resized to 128×128 and we apply histogram equalization to them to eliminate the inter-institute variations between the two datasets.

TABLE I

EVALUATION RESULTS ON BRATS DATASET

Method	Sample-level		Pixel-level	
	AUROC	AP	AUROC	AP
VQVAE [12]	0.66	0.82	0.75	0.22
ceVAE [7]	0.62	0.77	0.69	0.16
FPI [14]	0.70	0.84	0.71	0.32
Ours	**0.79**	**0.86**	**0.97**	**0.56**

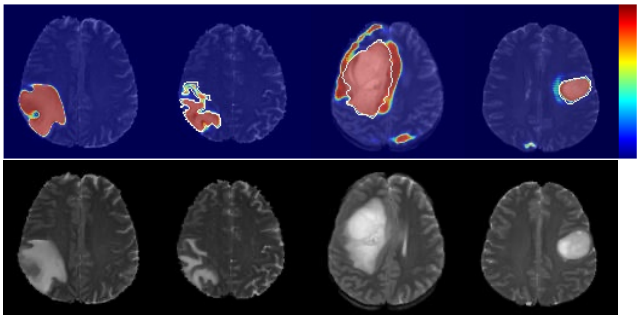

Fig. 3. Anomaly detection examples on BRATS MRIs. The anomalous regions are annotated with white contours; The red regions indicate significant abnormalities detected by our method.

Results and discussions: The quantitative results of anomaly detectio on the BRATS data set are reported in Table I. VQVAE [12] is an image reconstruction based method and reported as the state-of-the-art solution. Both ceVAE [7] and FPI [14] propose synthetic processing of the input data, so we re-implemented both methods for performance comparison. Fig. 3 shows several examples of anomaly detection by the proposed method. The white borders represent abnormality annotations provided in the BRATS set and red regions indicate detected abnormal regions with high anomaly scores by the proposed method.

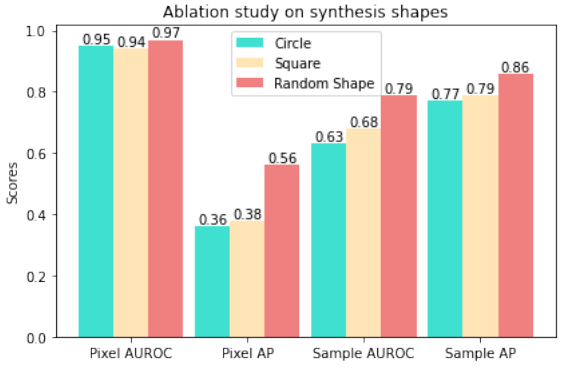

Fig. 4. Performance comparison with regular shape masking (Circle and Square) and random shape masking.

We further perform an ablation study on the effectiveness of shape randomness in pseudo-outlier synthesis. Specifically, we apply square masking from ceVAE [7], circular masking from FPI [14], and the proposed random-shape masking to the proposed solution in Fig. 1. As shown in 4, the random-shape pseudo-outlier synthesis strategy proposed in this paper significantly outperforms these regular shapes.

TABLE II

EVALUATION RESULTS ON MOOD DATASET

	Method	Sample-level		Pixel-level	
		AUROC	AP	AUROC	AP
Brain track	VQVAE [12]	**0.97**	0.92	**0.99**	0.81
	Ours	**0.97**	**0.95**	**0.99**	**0.91**
Abd. track	VQVAE [12]	0.83	0.73	**0.98**	0.57
	Ours	**0.93**	**0.87**	0.94	**0.66**

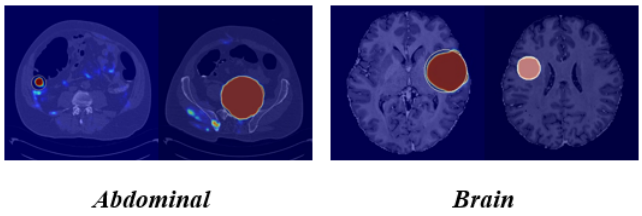

Fig. 5. Anomaly detection examples of the MOOD dataset. Please refer to Fig. 3 for label details.

C. Experiments on MOOD dataset

MOOD dataset: We follow the study in [12] and further evaluate our method on the Medical out-of-distribution (MOOD) analysis challenge dataset [19]. The MOOD set provides 800 normal Brain MRI voxels ($256 \times 256 \times 256$) in the task of brain anomaly detection and 550 normal abdominal CT voxels ($512 \times 512 \times 512$) in the abdominal track competition. The MOOD challenge provides 4 Brain MRI voxels and 4 Abdominal CT voxels as well as their abnormality ground truth for local, open evaluation.

Results and discussions: For the MOOD dataset, previous work VQVAE [12] has shown excellent performance, so we take it as our comparison baseline in this experiment. For a fair comparison, we follow their procedure to prepare the data and slice the MRI voxels into images. No further preprocessing is performed. The numerical results on MOOD images are reported in Table II, where the best performance values are highlighted. Quantitatively, our method significantly outperforms VQVAE [12] at both sample-level and pixel-level.

IV. CONCLUSIONS

We presented a novel self-supervised learning solution to anomaly detection. The introduced random-shape mask generation algorithm synthesized random-shape artificial outliers, which bridged the gap between irregular-shape abnormalities in natural MRI images and synthetic square/rectangular pseudo-anomalies in prior arts. With the synthetic dataset, we formulated the anomaly detection problem as a supervised anomaly scoring problem. Our experimentation demonstrated the superiority of the proposed methods on public medical imaging datasets.

REFERENCES

[1] B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Transactions on Medical Imaging, Institute of Electrical and Electronics Engineers (IEEE), 2014, pp.33.

[2] D. P. Kingma and M. Welling, "Auto-encoding variational bayes", International Conference on Learning Representations (ICLR), pp. 3, 2014.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative adversarial nets", Advances in neural information processing systems, pp. 2672-2680, 2014.

[4] X. Chen and K. EJapa, "Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders", 2018.

[5] T. Schlegl, P. Seeböck, S.M. Waldstein, U. Schmidt-Erfurth and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery", International conference on information processing in medical imaging, pp. 146-157, 2017.

[6] 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.

[7] D. Zimmerer, S. A. Kohl, J. Petersen, F. Isensee and K. H. Maier-Hein, "Context-encoding variational autoencoder for unsupervised anomaly detection", Proc. Med. Imag. Deep Learn., 2019.

[8] J. Tan, B. Hou, T. Day, J. Simpson, D. Rueckert, B. Kainz, "Detecting Outliers with Poisson Image Interpolation", Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., pp. 581-591, 2021.

[9] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. Huang, "Free-form image inpainting with gated convolution", Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 4471-4480, Oct. 2019.

[10] D. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, et al., "The human connectome project: a data acquisition perspective, " Neuroimage, 2012.

[11] D. Zimmerer et al., "Medical Out-of-Distribution Analysis Challenge." Zenodo, 2020, doi: 10.5281/zenodo.3715870.

[12] S. N. Marimont and G. Tarroni, "Anomaly Detection Through Latent Space Restoration Using Vector Quantized Variational Autoencoders," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1764-1767, doi: 10.1109/ISBI48211.2021.9433778.

[13] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell and A. A. Efros, "Context encoders: Feature learning by inpainting", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2536-2544, 2016.

[14] J. Tan, B. Hou, T. Day, J. Simpson, D. Rueckert, B. Kainz, "Detecting Outliers with Foreign Patch Interpolation", 2021.

[15] S. A. Taghanaki et al., "Combo loss: Handling input and output imbalance in multi-organ segmentation", Computerized Medical Imaging and Graphics, vol. 75, pp. 24-33, 2019.

[16] C.-L. Li, K. Sohn, J. Yoon and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization", arXiv:2104.04015, 2021.

[17] D. Zimmerer, F. Isensee, J. Petersen, S. Kohl and K. Maier-Hein, "Unsupervised anomaly localization using variational auto-encoders", Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., pp. 289-297, 2019.

[18] C. Baur, S. Denner, B. Wiestler, N. Navab and S. Albarqouni. "Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study", Medical Image Analysis, vol. 69, 2021.

[19] D. Zimmerer, J. Petersen, G. Köhler, P. Jäger, P. Full, T. Roß, T. Adler, A. Reinke, L. Maier-Hein and K. Maier-Hein, "Medical Out-of-Distribution Analysis Challenge 2021", 2021.