

# Spatial-aware Attention Generative Adversarial Network for Semi-supervised Anomaly Detection in Medical Image

Zerui Zhang<sup>1</sup>, Zhichao Sun<sup>1</sup>, Zelong Liu<sup>1</sup>, Zhou Zhao<sup>2</sup>, Rui Yu<sup>3</sup>, Bo Du<sup>1(✉)</sup>,  
and Yongchao Xu<sup>1(✉)</sup>

<sup>1</sup> National Engineering Research Center for Multimedia Software, School of Computer Science, Medical Artificial Intelligence Research Institute of Renmin Hospital, Wuhan University, Wuhan, China  
{dubo,yongchao.xu}@whu.edu.cn

<sup>2</sup> School of Computer Science, Central China Normal University, Wuhan, China

<sup>3</sup> University of Louisville, Louisville, KY, USA

**Abstract.** Medical anomaly detection is a critical research area aimed at recognizing abnormal images to aid in diagnosis. Most existing methods adopt synthetic anomalies and image restoration on normal samples to detect anomaly. The unlabeled data consisting of both normal and abnormal data is not well explored. We introduce a novel **Spatial-aware Attention Generative Adversarial Network (SAGAN)** for one-class semi-supervised generation of health images. Our core insight is the utilization of position encoding and attention to accurately focus on restoring abnormal regions and preserving normal regions. To fully utilize the unlabelled data, SAGAN relaxes the cyclic consistency requirement of the existing unpaired image-to-image conversion methods, and generates high-quality health images corresponding to unlabeled data, guided by the reconstruction of normal images and restoration of pseudo-anomaly images. Subsequently, the discrepancy between the generated healthy image and the original image is utilized as an anomaly score. Extensive experiments on three medical datasets demonstrate that the proposed SAGAN outperforms the state-of-the-art methods. Code is available at <https://github.com/zr728/SAGAN>

**Keywords:** Anomaly detection · Image restoration · Semi-supervised learning.

## 1 Introduction

The early anomaly detection of medical images plays a crucial role in the diagnosis and treatment of various pathologies. To alleviate the physician's reading burden and enhance diagnostic efficiency, computer-aided diagnosis (CAD) systems employing deep learning are gaining popularity [12]. However, obtaining annotations is consistently challenging in real-world medical datasets due to the

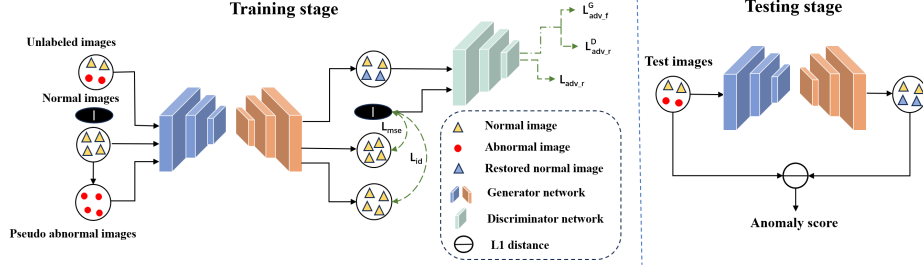
scarcity of anomalous data. This scarcity motivates the development of unsupervised anomaly detection (UAD), which operates without requiring annotations. Current UAD methods view anomaly detection as a one-class classification (OCC) problem [17]. In classical UAD approaches, only normal images are used for training to model the distribution of normal images. During the testing phase, out-of-distribution samples are identified as anomalies.

UAD methods can be roughly categorized into reconstruction-based, embedding-based, restoration-based and self-supervised-based methods. Reconstruction-based methods learn the latent space representation of normal data, and identify abnormal data by their reconstruction errors. These methods use autoencoder [6,8,29,13,2] or Generative Adversarial Networks (GANs) [20,19,1]. Embedding-based methods [18,16,7] primarily leverage discriminative features to model normal sample distributions in the feature space. Self-supervised-based methods propose synthetic anomalies to model normal distribution and detect abnormal structure [25,24,10,21]. Restoration-based methods [27,15,3,28,26] aim to restore pathological regions to healthy structures. However, the capacity to detect medical anomalies is significantly hindered by the absence of a large number of realistic anomaly images. Furthermore, a substantial amount of unlabeled data containing genuine anomalies exists within real-world medical datasets, yet these unlabeled images are overlooked by existing methods.

Recently, several approaches have highlighted the issue of underutilized unlabeled data [4,5,15,22]. The main concept behind Dual-distribution Discrepancy for Anomaly Detection (DDAD) [5] and Adaptation of pre-trained Masked AutoEncoder (AMAE) [4] involves training dual-distribution reconstruction networks to gauge the differences in reconstruction between abnormal or unlabeled data and normal data. However, these approaches typically necessitate two phases and multiple networks, making it more cumbersome. HealthyGAN [15] and Brainomaly [22] utilize normal images as supervision for generating healthy images with their unlabeled counterparts. Nevertheless, relying solely on normal image supervision is inadequate, as it may result in generating a healthy image that fails to restore abnormal regions on the original image, thus increasing the occurrence of false positives. Additionally, these methods overlook the recurrent anatomical structures present in most radiography images [28].

In this work, we propose a **S**patial-aware **a**ttention **G**enerative **A**dversarial **N**etwork (SAGAN) method for semi-supervised anomaly detection in medical images. SAGAN supervises the restoration of unlabeled data by ensuring accurate reconstruction of normal data and precise restoration of pseudo-anomaly regions. Recognizing the anatomical consistency inherent in most medical images, SAGAN divides the input image into  $N$  patches and incorporates positional conditional encoding for each patch. This strategy leverages the similarities found in structures at the same position, thereby enhancing the restoration quality.

Moreover, the primary challenge in unlabeled image restoration lies in accurately restoring the anomalous region rather than merely generating a healthy image, which might possess a structure entirely different from the original one. To address this issue, SAGAN incorporates an attention gate mechanism in-



**Fig. 1.** Overview of SAGAN. During the training stage, SAGAN learns to restore unlabeled images as normal images with the supervision of reconstructing normal images and restoring pseudo abnormal images. At the testing stage, the difference between the generated image and the original image reveals the presence of anomaly data.

spired by prior work [14,30], which dynamically emphasizes regions of interest and generates corresponding masks. These masks enable the SAGAN decoder to retain normal areas while effectively discarding abnormal regions of interest and generating the corresponding normal structures. Through extensive experimentation, we achieve state-of-the-art performance on three challenging medical datasets.

## 2 Methods

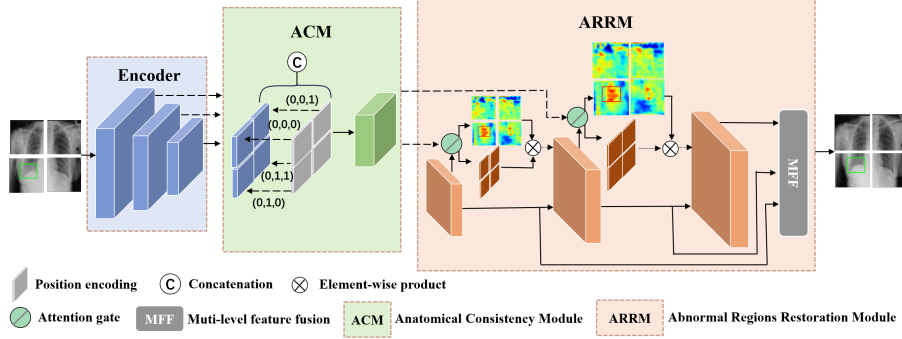
### 2.1 Overview

In the proposed SAGAN as shown in Fig. 1, we use a GAN-style framework to generate near-realistic healthy images from unlabeled images, healthy images and pseudo abnormal images. In concrete terms, the generator generates an additive map where each pixel represents the difference value of the input image that needs to be changed. Inspired by [28], we incorporate the use of a random binary mask during training phase to gate shortcut features, thereby enhancing feature aggregation. For each input image  $x$ , the final generated image  $x'$  following Eq. (1) and Eq.(2).

$$G(x) = I(x) \cdot \delta + (1 - \delta) \cdot x, \quad (1)$$

$$x' = \tanh(G(x) + x), \quad (2)$$

where  $G$  indicates generator function,  $\delta \sim \text{Bernoulli}(p)$  is a binary variable with  $p$  the gating probability.  $\tanh$  represents the  $\tanh$  activation function. Normal images  $x_n$ , pseudo anomaly images  $x_p$  and unlabeled images  $x_u$  are passed through the generator to get the corresponding healthy images  $x'_n, x'_p$  and  $x'_u$ .



**Fig. 2.** Generator of our proposed SAGAN. The main idea of the generator is to take advantage of the anatomical consistency of medical images and assign higher weights to the attention of anomaly features to refine the restoration of anomaly regions. Generator utilizes positional conditional encoding of patches to extract positional information and uses attention gate to adaptively learn the attention map of regions of interest to accurately control the restoration of anomaly regions.

We train SAGAN to ensure  $x'_u$  are close to reality with the supervision of reconstruction of  $x_n$  and restoration of  $x_p$ . We achieve the training objective by a identity loss, a reconstruction loss and an adversarial loss following Eq. (3), Eq. (4) and Eq. (5).

$$L_{id}^G = \mathbb{E}_{x_1 \in x_n} \|x'_1 - x_1\|_1, \quad (3)$$

$$L_{rec}^G = \mathbb{E}_{x_2 \in x_p, x_1 \in x_n} \|x'_2 - x_1\|_2, \quad (4)$$

$$L_{adv}^G = -\mathbb{E}_{x_3 \in x_u} [D(x'_3)], \quad (5)$$

$$L^G = L_{adv}^G + \lambda_{id} L_{id}^G + \lambda_{rec} L_{rec}^G, \quad (6)$$

Where  $x_2$  is the corresponding pseudo anomaly image generated from  $x_1$ .  $\lambda_{id}$  and  $\lambda_{rec}$  represent the weight of reconstruction loss and identify loss.  $D$  indicates the discriminator function to distinguish whether the generated image is real or fake. The discriminator classify normal images  $x'_n$  as real, as well as on healthy images  $x'_u$  generated from unlabeled images  $x_u$  as fake. The training objective is achieved using an adversarial loss with a gradient penalty following Eq. (7).

$$L_{adv}^D = \mathbb{E}_{x_3 \in x_u} [D(x'_3)] - \mathbb{E}_{x_1 \in x_n} [D(x_1)] + \lambda_{gp} \mathbb{E}_{\hat{x}} [(\|\nabla \hat{x} D(\hat{x})\|_2 - 1)^2], \quad (7)$$

where  $\hat{x}$  is a random weighted average of a batch of normal and generated healthy images, and  $\lambda_{gp}$  is the weight of gradient penalty to enhance training stability. In the above process, merely using a simple generator leads to inaccurate restoration. Accordingly, we design a spatial-aware generator consisting of anatomical

consistency module (ACM) and abnormal region restoration module (ARRM) to highlight abnormal regions and restore these regions to normal structures at the same position based on spatial information.

## 2.2 Positional condition encoding for anatomical consistency

To fully leverage the anatomical consistency of medical images, we divide the input image into non-overlapping patches of size  $N \times N$  and feed them into the generator’s encoder for feature extraction. To incorporate positional information and ensure anatomical consistency, we introduce an anatomical consistency module (ACM) adding positional condition encoding for each patch based on its position. Specifically, we initialize a binary code ranging from  $0 - N^2$ . The dimensions of the binary code is determined by  $\lceil \log_2(N \times N) + 1 \rceil$ . Taking  $N = 2$  as an example, the position of the first patch added is encoded as  $[0, 0, 0]$ . Then, the position encoding of patches are extended to corresponding feature. Both extracted features and skip-connections need to add position encoding. Subsequently, given an input image  $x$ , we refer to the resulting feature vectors as position encoding features  $f^{pe} \in \mathbb{R}^{B \times (C + \lceil \log_2(N \times N) + 1 \rceil) \times H \times W}$ . The spatially-aware features is transformed through the bottleneck layer into  $f^s \in \mathbb{R}^{B \times C \times H \times W}$ .

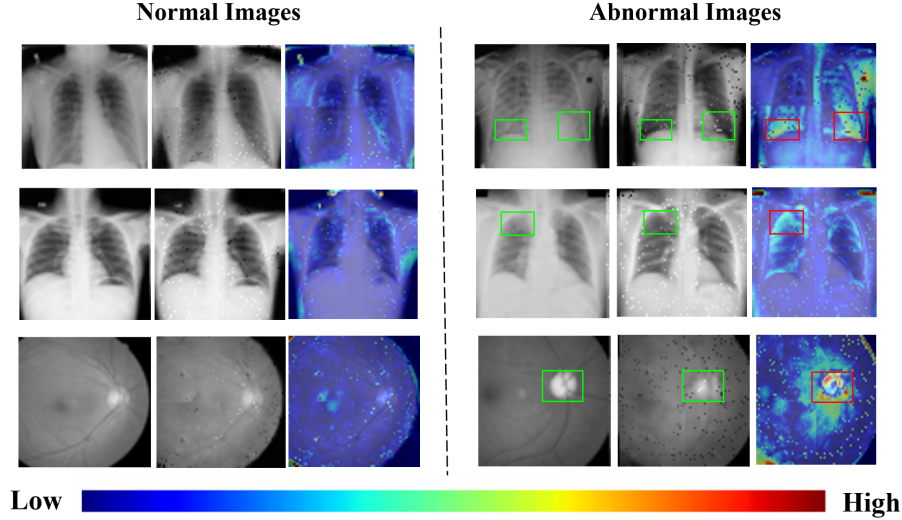
## 2.3 Abnormal region restoration module

The most significant challenge in the restoration task is accurately restoring abnormal regions. We define this process as preserving the normal region, ‘dropping’ the abnormal region and generating a corresponding normal one to replace it. We introduce attention gate to each level of the decoder to achieve this objective. The most important part of attention gate is the attention coefficient,  $\alpha_i \in [0, 1]$ , which denotes the scalar attention value for each pixel to identify salient feature and to highlight the abnormal activation. Due to information extracted from coarse scale is used in gating to disambiguate irrelevant and noisy responses in skip connections [14], we choose low-level skip-connections as gating vector  $g^l \in \mathbb{R}^{B \times C^l \times H^l \times W^l}$ , where  $l$  is the number of skip-connection used in training,  $C$ ,  $H$  and  $W$  denote the channel, height and width of  $l^{th}$  skip-connection. Low-level skip-connections highlight anomaly features by increasing the value of the attention coefficient of the anomaly region and decreasing the value of attention coefficient in normal regions. The calculation process of attention gate can be expressed as:

$$\alpha = S(C_3(R(C_1(f^l) + C_2(g^l)))), \quad (8)$$

$$g^l = \alpha \times g^l, \quad (9)$$

where  $f^l$  represent feature of the input of the  $l^{th}$  layer of the decoder,  $C_1$ ,  $C_2$  and  $C_3$  denote convolution layers learning attention coefficients to assign weights to different positions of the parallel input signal.  $R$  and  $S$  represent ReLU function and sigmoid function. After being filtered through attention gates and decoder, the feature maps are upsampled and fused together using an Multi-level feature fusion module to generate the final restored image with high abnormal response.



**Fig. 3.** Visualization of heatmaps on medical datasets. The heatmap is derived from the difference maps between the restored image and the original image. Images from top to bottom are from RSNA, VinDr-CXR and LAG datasets, respectively. Original images, restored images, and heatmap visualizations are arranged from left to right. Green boxes represent abnormal regions and red boxes indicate the corresponding abnormal regions on heatmaps.

### 3 Experiments

#### 3.1 Datasets and implementation details

**Datasets.** We evaluate our method on two CXR datasets and one retinal fundus image dataset: 1) RSNA Pneumonia Detection Challenge dataset<sup>1</sup>; 2) VinBig-Data Chest X-ray Abnormalities Detection Challenge dataset (VinDr-CXR)<sup>2</sup>; 3) Large-scale Attention-based Glaucoma (LAG) dataset [11]. We show a summary of each dataset’s repartitions in Table 1. We utilize both normal images  $T_n$  and unlabeled images  $T_u$  for model training.

**Implementation details.** The discriminator network follows PatchGAN [9,31] architecture and is similar to the ones used in [23]. In training stage, we update the weights of the generator once for every two weight updates of the discriminator network. We employ a simple approach to transform a normal image into a corresponding pseudo anomaly image referenced to FPI [24]. We adopt Adam optimizer and set the learning rate and batch size to  $5e^{-5}$  and 64, respectively. The learning rate has been decayed every 1000 iterations. The maximum epoch is set to 100000 across all settings. For a fair comparison, all the input images

<sup>1</sup> <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>

<sup>2</sup> <https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection>

**Table 1.** Summary of dataset repartitions. Unlabeled image set  $T_u$  is constructed from the images presented in parentheses without any annotation.

| Dataset   | Normal training set $T_n$ | Unlabeled training set $T_u$       | Testing set $T_{test}$      |
|-----------|---------------------------|------------------------------------|-----------------------------|
| RSNA      | 3851                      | 3851 (4000 normal + 5012 abnormal) | 1000 normal + 1000 abnormal |
| VinDr-CXR | 4000                      | 4000 (5606 normal + 3394 abnormal) | 1000 normal + 1000 abnormal |
| LAG       | 1500                      | 1500 (832 normal + 900 abnormal)   | 811 normal + 811 abnormal   |

**Table 2.** Performance comparison with SOTA methods. We compare our proposed SAGAN with these state-of-the-art methods which utilize normal images and unlabeled images. We conduct three experiments on each dataset and choose the average value as the final results of the SAGAN. The best results in each category are **bolded** and the second best results are underlined.

| Method          | Taxonomy       | VinDr-CXR    |              | RSNA         |              | LAG          |              |
|-----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                 |                | AP(%)        | AUC(%)       | AP(%)        | AUC(%)       | AP(%)        | AUC(%)       |
| CutPaste [10]   | Self-sup.      | 60.0         | 59.2         | 61.7         | 59.8         | 51.7         | 48.9         |
| FPI [24]        | Self-sup.      | 49.4         | 47.4         | 53.8         | 46.6         | 56.1         | 52.9         |
| PII [25]        | Self-sup.      | 67.2         | 66.8         | 85.4         | 84.3         | 63.1         | 63.1         |
| NSA [21]        | Self-sup.      | 64.8         | 64.4         | 84.3         | 84.2         | 68.0         | 68.6         |
| DDAD [5]        | Rec.+ Self-sup | 84.3         | 85.9         | 91.6         | 91.3         | 92.3         | 93.1         |
| AMAE [4]        | Rec.           | <u>84.5</u>  | 86.1         | <u>91.7</u>  | <u>91.4</u>  | -            | -            |
| Brainomaly [22] | Restoration.   | 84.37        | <u>88.54</u> | 87.93        | 91.01        | <u>94.71</u> | <u>95.92</u> |
| <b>SAGAN</b>    | Restoration.   | <b>89.30</b> | <b>91.47</b> | <b>92.60</b> | <b>92.45</b> | <b>96.96</b> | <b>96.98</b> |

are resized to  $64 \times 64$  pixels. All experiments are implemented using PyTorch and train each models on a single GeForce RTX 2080 Ti GPU. We use the area under the ROC curve (AUC) and average precision (AP) as the evaluation metrics.

### 3.2 Main results

In Table 2, we compare our proposed method with a wide range of state-of-the-art (SOTA) methods, including self-supervised synthetic anomaly. We use the unified implementation provided by NSA [21] for CutPaste, FPI and PII. All other methods used in the experiments are implemented using their official codes. On the RSNA dataset, SAGAN outperforms the previous SOTA method AMAE [4] and the baseline method Brainomaly [22] by achieving gains of 1.05%, 1.00% and 4.67%, 1.44% in AP, AUC. On the VinDr-CXR dataset, SAGAN surpass the AMAE by 4.80%, 5.37% in AP, AUC, respectively, and outperforms Brainomaly by 4.93%, 2.93%. On the LAG dataset, SAGAN outperforms the previous SOTA method DDAD [5] and the baseline method by 4.66%, 3.88% and 2.25%, 1.06% in AP, AUC. We show qualitative results from RSNA dataset, VinDr-CXR dataset and LAG datasets in Fig. 3. Our SAGAN maintains the content of normal images unchanged, while restoring abnormal regions in abnormal images.

**Table 3.** Ablation study on VinDr-CXR dataset [11] for each component of the SAGAN. ACM and ARRM represent Anatomical Consistency Module and Abnormal Regions Restoration Module, respectively.

| Method         | AP (%)↑      | AUC (%)↑     |
|----------------|--------------|--------------|
| Baseline [22]  | 84.37        | 88.54        |
| + ACM          | 86.07        | 89.89        |
| + ARRM         | 87.57        | 90.37        |
| + ACM and ARRM | <b>89.30</b> | <b>91.47</b> |

**Table 4.** Ablation study on RSNA dataset with a varying anomaly ratio (AR) of  $T_u$ . AR =  $0_u$  denotes experiments trained without unlabeled data.

| AR    | AP (%)↑ | AUC (%)↑ |
|-------|---------|----------|
| $0_u$ | 65.87   | 70.54    |
| 0     | 75.99   | 77.75    |
| 0.2   | 89.65   | 89.23    |
| 0.4   | 92.14   | 89.89    |
| 0.6   | 92.60   | 92.45    |
| 0.8   | 92.64   | 92.42    |

### 3.3 Ablation studies

**Quantitative analysis of spatial-aware attention.** To understand the effectiveness of each components of SAGAN. We conduct ablation experiments on the VinDr-CXR dataset presented in Table 3. It can be clearly seen that the baseline method Brainomaly [22] achieve 84.37% AP. Adding the proposed anatomical consistency module, the AP increases by 1.70%. On the other, abnormal regions restoration module improved AP by 3.20% suggesting that this module effectively highlights anomaly features. Furthermore, with the addition of these two modules, SAGAN achieved a 4.93% improvement in AP. Accordingly, the results of these experiments demonstrate the effectiveness of each components of SAGAN.

**Quantitative analysis of anomaly ratio of unlabeled images.** In real-world clinical settings, the anomaly ratio (AR) within vast amounts of unlabeled images is unknown. To simulate the situation, we conduct ablation experiments on the RSNA dataset, varying the AR of  $T_u$  from 0% to 100%, as detailed in Table 3. In an extreme experiment, we simulate the absence of unlabeled images by treating pseudo anomaly images as fake in the discriminator. Remarkably, even with an AR of 0%, the SAGAN consistently outperforms this scenario, showcasing its ability to enhance restoration performance across all AR situations. Notably, even when the AR of unlabeled data is 20%, the SAGAN substantially improves performance, highlighting its capacity to leverage unlabeled images containing anomalies for significant clinical advancements. To ensure fairness in comparison, all experiments are conducted under AR = 60%.

## 4 Conclusion

In this paper, we propose the Spatial-aware Attention Generative Adversarial Network for anomaly detection (SAGAN), which fully utilizes anomaly spatial features in unlabeled images to efficiently restore anomaly regions. Experiments

on three benchmarks demonstrate that SAGAN achieves state-of-the-art performance. As for the limitation, SAGAN is not able to localize anomalies at the pixel level precisely and SAGAN still requires a generous number of normal annotations in training stage. In the future work, we will extend our SAGAN method to achieve accurate anomaly localization and explore the potential of anomaly detection without any training annotations.

**Acknowledgments.** This work was supported in part by the National Key Research and Development Program of China (2023YFC2705700), NSFC 62222112, 62225113, and 62176186, the Postdoctoral Fellowship Program of CPSF (GZC20230924), the Innovative Research Group Project of Hubei Province under Grants (2024AFA017), the NSF of Hubei Province of China (2024AFB245), and CAAI Huawei MindSpore Open Fund.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Asian Conf. Comput. Vis. pp. 622–637. Springer (2019)
2. Bercea, C.I., Rueckert, D., Schnabel, J.A.: What do aes learn? challenging common assumptions in unsupervised anomaly detection. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 304–314. Springer (2023)
3. Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.: Reversing the abnormal: Pseudo-healthy generative networks for anomaly detection. Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. (2023)
4. Bozorgtabar, B., Mahapatra, D., Thiran, J.P.: Amae: Adaptation of pre-trained masked autoencoder for dual-distribution anomaly detection in chest x-rays. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 195–205. Springer (2023)
5. Cai, Y., Chen, H., Yang, X., Zhou, Y., Cheng, K.T.: Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. *medical image analysis* **86**, 102794 (2023)
6. Chen, Y., Tian, Y., Pang, G., Carneiro, G.: Deep one-class classification via interpolated gaussian descriptor. In: Proc. AAAI Conf. Artif. Intell. vol. 36, pp. 383–392 (2022)
7. Dong, H., Zhang, Y., Gu, H., Konz, N., Zhang, Y., Mazurowski, M.A.: Swssl: Sliding window-based self-supervised learning for anomaly detection in high-resolution images. *IEEE Trans. Medical Imaging*. (2023)
8. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Int. Conf. Comput. Vis. pp. 1705–1714 (2019)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 1125–1134 (2017)

10. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 9664–9674 (2021)
11. Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: Attention based glaucoma detection: A large-scale database and cnn model. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 10571–10580 (2019)
12. Luo, L., Yu, L., Chen, H., Liu, Q., Wang, X., Xu, J., Heng, P.A.: Deep mining external imperfect data for chest x-ray disease screening. *IEEE Trans. Medical Imaging*. **39**(11), 3583–3594 (2020)
13. Naval Marimont, S., Tarroni, G.: Implicit field learning for unsupervised anomaly detection in medical images. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 189–198. Springer (2021)
14. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
15. Rahman Siddiquee, M.M., Shah, J., Wu, T., Chong, C., Schwedt, T., Li, B.: Healthygan: Learning from unannotated medical images to detect anomalies associated with human disease. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 43–54. Springer (2022)
16. Reiss, T., Cohen, N., Bergman, L., Hoshen, Y.: Panda: Adapting pretrained features for anomaly detection and segmentation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 2806–2814 (2021)
17. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: Proc. Int. Conf. Mach. Learn. pp. 4393–4402. PMLR (2018)
18. Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M.H., Rabiee, H.R.: Multiresolution knowledge distillation for anomaly detection. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 14902–14912 (2021)
19. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *medical image analysis* **54**, 30–44 (2019)
20. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 146–157. Springer (2017)
21. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for self-supervised anomaly detection and localization. In: Eur. Conf. Comput. Vis. pp. 474–489. Springer (2022)
22. Siddiquee, M.M.R., Shah, J., Wu, T., Chong, C., Schwedt, T.J., Dumkrieger, G., Nikolova, S., Li, B.: Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images. In: WACV. pp. 7573–7582 (2024)
23. Siddiquee, M.M.R., Zhou, Z., Tajbakhsh, N., Feng, R., Gotway, M.B., Bengio, Y., Liang, J.: Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In: Int. Conf. Comput. Vis. pp. 191–200 (2019)
24. Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., et al.: Detecting outliers with foreign patch interpolation. *Machine Learning for Biomedical Imaging* **1**, 1–27 (2022)
25. Tan, J., Hou, B., Day, T., Simpson, J., Rueckert, D., Kainz, B.: Detecting outliers with poisson image interpolation. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 581–591 (2021)

26. Tang, Y., Tang, Y., Zhu, Y., Xiao, J., Summers, R.M.: A disentangled generative model for disease decomposition in chest x-rays via normal image synthesis. *medical image analysis* **67**, 101839 (2021)
27. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. pp. 35–45. Springer (2022)
28. Xiang, T., Zhang, Y., Lu, Y., Yuille, A.L., Zhang, C., Cai, W., Zhou, Z.: Squid: Deep feature in-painting for unsupervised anomaly detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.* pp. 23890–23901 (2023)
29. Zhao, H., Li, Y., He, N., Ma, K., Fang, L., Li, H., Zheng, Y.: Anomaly detection for medical images using self-supervised and translation-consistent features. *IEEE Trans. Medical Imaging*. **40**(12), 3641–3651 (2021)
30. Zhao, W., Xu, L.: Weakly supervised target detection based on spatial attention. *Visual Intelligence* **2**(1), 2 (2024)
31. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Int. Conf. Comput. Vis.* pp. 2223–2232 (2017)