MICCAI

Position-Guided Prompt Learning for Anomaly Detection in Chest X-Rays

Zhichao Sun[1], Yuliang Gu[1], Yepeng Liu[1], Zerui Zhang[1], Zhou Zhao[2], and Yongchao Xu[1]

[1] National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence, School of Computer Science, Medical Artificial Intelligence Research Institute of Renmin Hospital, Wuhan University, Wuhan, China
`yongchao.xu@whu.edu.cn`
[2] School of Computer Science, Central China Normal University, Wuhan, China

Abstract. Anomaly detection in chest X-rays is a critical task. Most methods mainly model the distribution of normal images, and then regard significant deviation from normal distribution as anomaly. Recently, CLIP-based methods, pre-trained on a large number of medical images, have shown impressive performance on zero/few-shot downstream tasks. In this paper, we aim to explore the potential of CLIP-based methods for anomaly detection in chest X-rays. Considering the discrepancy between the CLIP pre-training data and the task-specific data, we propose a position-guided prompt learning method. Specifically, inspired by the fact that experts diagnose chest X-rays by carefully examining distinct lung regions, we propose learnable position-guided text and image prompts to adapt the task data to the frozen pre-trained CLIP-based model. To enhance the model's discriminative capability, we propose a novel structure-preserving anomaly synthesis method within chest x-rays during the training process. Extensive experiments on three datasets demonstrate that our proposed method outperforms some state-of-the-art methods. The code of our implementation is available at https://github.com/sunzc-sunny/PPAD.

Keywords: Anomaly detection · Chest X-ray · Prompt learning.

1 Introduction

Chest X-ray remains the most frequently utilized diagnostic tool for detecting lung diseases. Thus, the rise of automated chest X-ray analysis through deep learning techniques has gained considerable attention. However, pathology annotations of chest X-rays are hard to obtain, requiring expert experience. Consequently, unsupervised anomaly detection methods [28,27] are proposed, aiming to identify diseases in the testing set using only normal data from the training set. Previous methods [25,12,22,2,1,3] mainly attempt to model the distribution of normal images during training, and then identify the samples that do not conform to normal profile as anomalies.

Recently, Constrastive Language-Image Pre-training (CLIP) [15] has gained increasing attention in the field of medical imaging. Advancements [21,23,4] demonstrate the effectiveness of CLIP-based methods in downstream tasks when pre-trained on large-scale chest X-ray datasets with image-text pairs. Specifically, CheXzero [21] pre-trains CLIP on a large chest X-ray dataset for zero-shot classification. KoBo [4] integrates clinical knowledge into the pre-training process, achieving comparable results in eight downstream tasks. Despite the effectiveness of CLIP for downstream tasks, its massive parameters and extensive training data requirement make it impractical to fine-tune the full model for specific tasks. To address this challenge, prompt engineering and prompt learning methods [14,26,10,24] have been proposed. For instance, Xplainer [14] focuses on designing appropriate expert-level text input in prompt engineering. CoOp [26] proposes learnable text contexts for prompt learning. MaPLe [10] presents a multi-modal prompt learning framework for both vision and language branches of CLIP. However, these methods do not directly consider the domain differences between pre-training data and task-specific data. Besides, they do not take into account the positional prior in chest X-rays.

In this work, we propose a **P**osition-guided **P**rompt learning method for **A**nomaly **D**etection in chest X-rays (**PPAD**). PPAD leverages learnable text prompt and image prompt to minimize the gap between pre-training data and task-specific data. The learnable prompts are inserted into the inputs before being fed into the encoders. Besides, inspired by the fact that experts diagnose chest X-rays by carefully examining distinct lung regions, we leverage the positional prior in chest X-rays to introduce position-guided prompts. Through the position-guided prompts, the model can focus on various regions, simulating the diagnostic process of experts.

Furthermore, we propose a **S**tructure-preserving **A**nomaly **S**ynthesis method (**SAS**) in the training phase to cope with the issue that learning solely from normal samples may limit model's discrimination capability. Previous anomaly synthesis methods [19,11,20,18] are primarily focused on cutting from one normal regular region and pasting to another to create synthetic anomaly. However, these methods distort organ structures, potentially leading to model overfitting to this scenario. On the contrary, we utilize Gamma correction on the input image within a random irregular mask to simulate lesions while preserving the structure of lung. The gamma value in Gamma correction associates with distance to mask's boundary, which allows for seamless anomaly insertion. Experimental results indicate that SAS generates more authentic anomalies, which is beneficial for anomaly detection in chest X-rays. Besides, the proposed PPAD obtains state-of-the-art performance on three chest X-ray anomaly detection datasets.

2   Method

An overview of the proposed PPAD is depicted in Fig. 1. We employ position-guided prompts to adapt both the text and image inputs before feeding them into the encoders. The classification word of text input is substituted with a class
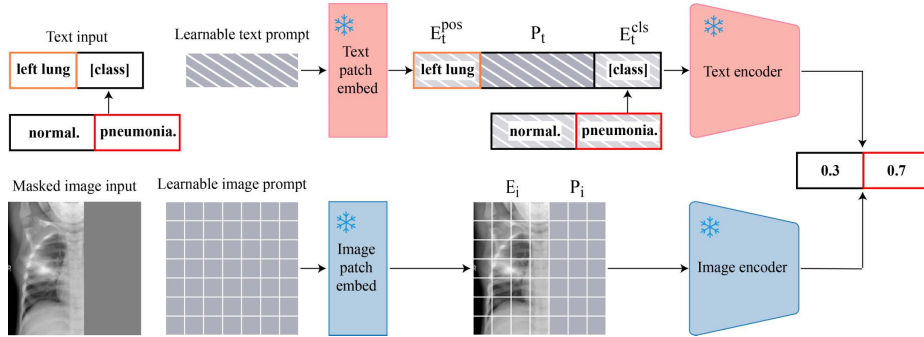
Fig. 1. The pipeline of our proposed PPAD. The main idea of the PPAD is to adapt text data and image data using learnable prompts. The rest of the model is frozen. Four positional prompts are optional. Taking the "left lung" for example, PPAD incorporates "left lung" as the position prompt embeddings E_t^{pos}. The learnable text prompt P_t is insert between position prompt embeddings E_t^{pos} and class embeddings E_t^{cls}. Image input is either the synthetic anomaly or the normal image. The right lung region of the input image embedding E_i is replaced by the learnable image prompt P_i.

name, either "normal" or "pneumonia". Subsequently, text features and image features are respectively extracted through the text encoder and image encoder. Finally, the prediction probability is computed by applying cosine similarity between the text and image features, followed by a softmax function.

2.1 Position-guided prompt learning.

We leverage the position prior of the chest X-ray to introduce a Position-guided Prompt learning method for chest X-ray Anomaly Detection (PPAD). Specifically, we divide the image into four sections: left lung, right lung, upper lung, and lower lung. Taking the left lung as an example (see Fig. 1), we initiate the text input with "left lung". For the image input, we focus on the left lung within the image, masking the right lung with a binary mask M. We mask the right lung to prevent it from influencing the result. Subsequently, we introduce learnable text prompt P_t and image prompt P_i. The text input is converted into token embeddings by the text patch embed module. We separate it into position prompt embeddings E_t^{pos} and class embeddings E_t^{cls}. The learnable text prompt is inserted between E_t^{pos} and E_t^{cls}. The text input E_{text} of the text encoder emerges as a combination, denoted as:

$$E_{text} = E_t^{pos} \oplus P_t \oplus E_t^{cls}, \tag{1}$$

where $\oplus$ symbolizes embedding concatenation. The masked image is divided into non-overlapping patches and is further tokenized into embeddings E_i, by the image patch embed module. We replace the image patches within the mask M with the learnable image prompt patches. The image input E_{image} of the
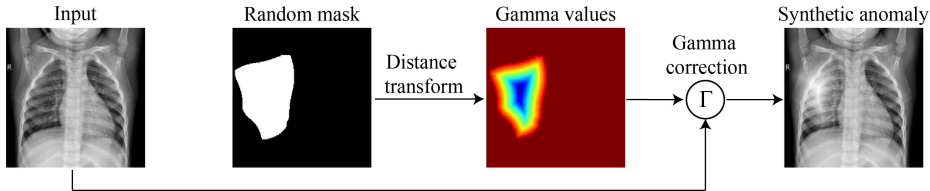
Fig. 2. The illustration of the proposed SAS. SAS applies distance transform within a random mask to create smoothed gamma values. Then, synthetic anomaly is generated via Gamma correction applied to the input normal image.

image encoder is computed as:

$$E_{image} = E_i \odot M + P_i \odot (1 - M), \tag{2}$$

where $\odot$ symbolizes element-wise multiplication. The learnable prompts have the capacity to adapt the training data to the frozen pre-trained CLIP-based model. Besides, employing positional guidance allows the model to focus on the guided region, simulating the diagnosis results obtained by experts who observe lesions in various regions.

2.2 Structure-preserving anomaly synthesis

We propose a Structure-preserving Anomaly Synthesis method (SAS) during the training phase to enhance the model's discriminative capability. The process is depicted in Fig. 2. Firstly, we generate a random, irregular anomaly mask M_a through several steps: 1) Randomly select ten points following the Perlin noise distribution; 2) Determine the smallest enclosing convex polygon for these points; 3) Randomly replace the edges of polygon with second-order Bézier curves; 4) Fill the interior to achieve the random irregular mask. After generating the anomaly mask, we utilize a distance transform to calculate the distance $D(x)$ from each point x within the anomaly mask to its nearest boundary point. Then, smoothed gamma values $\gamma(x)$ on each point x within the anomaly mask M_a are given by the randomly assigned weight w, using the formula:

$$\gamma(x) = 1 + \frac{D(x)}{\max\limits_{x' \in M_a} D(x')} * w, \tag{3}$$

where $w > -1$ is a constraint to ensure $\gamma > 0$. Outside the anomaly mask, γ is set to 1. Finally, the synthetic anomaly results from employing gamma correction $\Gamma(I) = I^\gamma$ on the input image I. Inside the mask, the parameter γ is dependent on the distance D, ensuring that the closer to the mask's boundary, the less pronounced the grayscale value changes. Outside the mask, grayscale remains unchanged. The distance-related gamma correction allows for seamless anomaly insertion and preserves the lung structure. Additionally, we generate irregular masks, rather than rectangle masks used in previous methods, to produce more authentic anomalies.

Table 1. Benchmark results on ZhangLab dataset [9] and CheXpert dataset [7]. Top group: results of previous SOTA methods. Bottom group: results of CLIP-based methods and prompt learning methods. The best results in each group are denoted with **bold**. Methods marked with an asterisk (*) are reproduced in this work. Other results are either from their original paper or from SQUID [22]. Mean and standard deviation are the results of five independent experiments.

Method	ZhangLab dataset [9]			CheXpert dataset [7]		
	ACC (%)↑	AUC (%)↑	F1 (%)↑	ACC (%)↑	AUC (%)↑	F1 (%)↑
MemAE [6]	56.5±1.1	77.8±1.4	82.6±0.9	55.6±1.4	54.3±4.0	53.3±7.0
SALAD [25]	75.9±0.9	82.7±0.8	82.1±0.3	-	-	-
CutPaste [11]	64.0±6.5	73.9±3.9	72.3±8.9	62.7±2.0	65.5±2.2	60.3±4.6
PANDA [16]	65.4±1.9	65.7±1.3	66.3±1.2	66.4±2.8	68.6±0.9	65.3±1.5
M-KD [17]	69.1±0.2	74.1±2.6	62.3±8.4	66.0±2.5	69.8±1.6	63.6±5.7
IF 2D [12]	76.4±0.2	81.0±2.8	82.2±2.7	-	-	-
SQUID [22]	**80.3±1.3**	**87.6±1.5**	**84.7±0.8**	**71.9±3.8**	**78.1±5.1**	**75.9±5.7**
CheXzero* [21]	83.0	92.7	87.5	77.4	87.7	79.7
Xplainer* [14]	78.2	89.9	85.0	75.6	83.6	75.6
CoOp* [26]	84.6±1.7	94.6±1.1	88.6±1.0	80.4±1.8	87.0±1.3	79.2±2.2
MaPLe* [10]	86.1±1.0	95.1±1.2	89.5±0.5	79.5±0.8	86.3±1.0	79.6±1.2
PPAD (Ours)	**89.4±0.6**	**96.7±0.4**	**91.8±0.5**	**82.7±0.6**	**88.5±0.9**	**82.0±0.8**

2.3 Training and inference

During the training process, we choose a random position prefix, which can be any one among "left lung", "right lung", "upper lung", "lower lung", or "- -" (- - symbolizes empty). The empty position prefix indicates a view of the entire lung, where the position name is null and the image is unmasked. We then apply SAS to generate synthetic anomaly with a 50% probability. We employ the standard Binary Cross-Entropy (BCE) loss as the loss function. It is important to note that we only optimize the learnable text prompt and image prompt, keeping the other pre-trained parameters frozen.

During the inference phase, we evaluate each image at four separate positions and the entire view, obtaining five distinct probabilities. If the maximum value among these probabilities exceeds a threshold η, we employ this maximum value for classification. Otherwise, we calculate the average of these probabilities as the final outcome.

3 Experiments

3.1 Datasets and implementation details

Datasets. We evaluate our method on three public chest X-ray datasets: 1) ZhangLab Chest X-ray dataset[3] [9], 2) Stanford CheXpert dataset[4] [7], 3) Vin-

---

[3] https://data.mendeley.com/datasets/rscbjbr9sj/3
[4] https://stanfordmlgroup.github.io/competitions/chexpert

Table 2. Benchmark results on VinDr-CXR dataset [13]. The results of top group are either from their respective original paper or from DDAD [3].

| Method	CutPaste [11]	IGD [5]	AMAE [2]	Baugh *et al.* [1]	DDAD [3]
AUC (%)↑	70.2	59.2	74.2	76.6	**78.2**
AP (%)↑	69.8	58.7	72.9	**78.4**	74.6
Method	CheXzero* [21]	Xplainer* [14]	CoOp* [26]	MaPLe* [10]	PPAD (Ours)
AUC (%)↑	77.6	76.5	79.9±2.1	74.2±1.3	**81.9±0.9**
AP (%)↑	78.4	78.2	80.9±2.1	75.7±1.1	**82.1±1.1**

BigData Chest X-ray Abnormalities Detection Challenge dataset (VinDr-CXR)[5] [13]. We employ the dataset setup and evaluation metrics from SQUID [22] for the ZhangLab and CheXpert datasets, while those from DDAD [3] are used with the VinDr-CXR dataset. It is important to note that both Xplainer [14] and our baseline method (CheXzero [21]), are pre-trained on the MIMIC-CXR dataset [8], which is unrelated to the datasets utilized in our task. Details of datasets setup are described in the Supplementary Material.

Implementation details. We implement the proposed PPAD, together with CoOp [26] and MaPLe [10], based on CheXzero [21] with the proposed SAS employed as the anomaly synthesis method. The total training epoch is set to 100. By default, the hyperparameter w involved in Eq. 3 is randomly chosen from {-0.999, -0.99, 2, 3} to introduce diverse synthetic anomalies. The threshold η is set to 0.8. The few-shot evaluation protocol established in CLIP [15] is adopted, which uses 64 shots randomly selected from normal training sets during training. We report the mean and standard deviation of the proposed PPAD and the reproduced methods, computed over five runs.

3.2   Main results

Table 1 and table 2 are divided into two groups for comparison. The first group compares PPAD with some SOTA methods [6,25,11,16,17,12,22,5,2,1,3]. The second group, PPAD compares with both CLIP-based methods [21,14] and prompt learning methods [26,10]. On the ZhangLab dataset [9], PPAD outperforms previous SOTA method SQUID [22] by 9.1%, 9.1%, 7.1% in ACC, AUC, and F1, respectively. When compared to the CLIP-based methods CheXzero [21] and Xplainer [14], PPAD achieves 6.4%, 4.0%, 4.3% and 11.2%, 6.8%, 6.8% in ACC, AUC, and F1 improvement. Furthermore, in comparison to the prompt learning methods CoOp [26] and MaPLe [10], PPAD achieves improvements of 4.8%, 2.1%, 3.2% and of 3.3%, 1.6%, 2.3% in ACC, AUC, and F1, respectively. The same conclusion can be drawn on the other two datasets. On the CheXpert dataset [7], we surpass SQUID by 10.8%, 10.4%, and 6.1% in ACC, AUC, F1,

---

[5] https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection
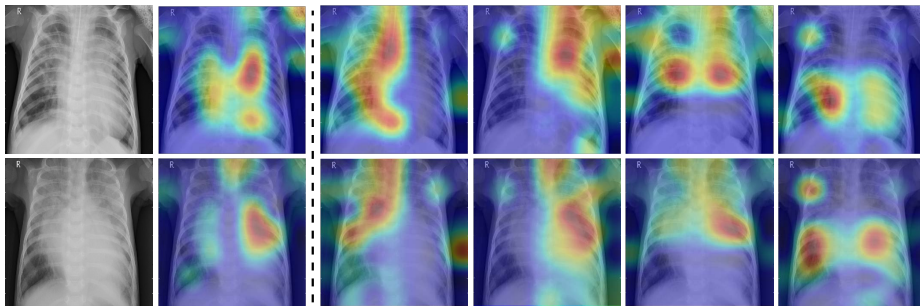
(a) Input   (b) Entire lung   (c) Left lung   (d) Right lung   (e) Upper lung   (f) Lower lung

Fig. 3. The visualization of CAM guided by the entire view (the second column) and various position prompts (last four columns).

Table 3. Ablation study on Zhanglab dataset [9] for the contribution of each position-guided prompt in PPAD.

Method	ACC (%)↑	AUC (%)↑	F1 (%)↑	AP (%)↑
Baseline (CheXzero [21])	83.0	92.7	87.5	95.4
+ Learnable text prompt (CoOp [26])	84.6±1.7	94.6±1.1	88.6±1.0	96.7±0.9
+ Position-guided text prompt	86.7±2.2	95.9±0.5	89.3±2.5	97.7±0.3
+ Position-guided text and image prompt	**89.4±0.6**	**96.7±0.4**	**91.8±0.5**	**98.0±0.3**

respectively, and outperform CheXzero by 5.3%, 0.8%, and 2.3%. On the VinDr-CXR dataset [13], PPAD outperforms DDAD [3] and CheXzero by 3.7%, 7.5% and 4.3%, 3.7% in AUC, AP, respectively.

3.3   Ablation studies

Visualization of position-guided impact. The impact of position guidance is visualized through the Class Activation Mapping (CAM) in Fig. 3. In the first row, the entire lung view fails to detect anomalies in the left region. However, specific views focused on the left, upper, and lower sections of the lung successfully identify these anomalies. In the second row, the left lung view detects anomalies that are missed in the entire lung view. These outcomes indicate the importance of utilizing diverse position prompts.

Quantitative analysis of position-guided prompts. As presented in Table 3, the baseline method CheXzero [21] achieves an ACC of 83.0%. The introduction of a learnable text prompt (CoOp [26]) improves the ACC by 1.6%. Introducing the position-guided text prompt further improves the ACC by 2.1%. Subsequently, the integration of the position-guided learnable image prompt leads to an extra increase of 2.7%. These results demonstrate the contribution of each position-guided prompt in PPAD.

Table 4. Ablation study on applying various anomaly synthesis methods to the PPAD.

ZhangLab [9]	ACC (%)↑	AUC (%)↑	F1(%)↑	AP(%)↑
PPAD + FPI [19]	87.0±0.8	94.2±0.7	90.0±0.8	96.2±0.7
PPAD + PII [20]	88.1±1.1	95.7±0.7	90.5±1.2	97.5±0.4
PPAD + CutPaste [11]	87.0±0.8	94.4±0.8	89.7±0.6	96.4±0.6
PPAD + AnatPaste [18]	87.3±1.1	95.4±1.3	90.0±0.7	97.3±0.8
PPAD + SAS (Ours)	**89.4±0.6**	**96.7±0.4**	**91.8±0.5**	**98.0±0.3**

Quantitative analysis of the proposed SAS. As detailed in Table 4, the proposed SAS outperforms other anomaly synthesis methods including FPI [19], PII [20], CutPaste [11], and AnatPaste [18], all of which are based on the PPAD framework. The SAS synthesises anomalies that resemble natural diseases, thereby improving the detection of anomalies in chest X-rays. The visualization results of different synthetic anomalies are shown in Supplementary Material.

Ablation with varied numbers of training shots. We perform an ablation study focusing on the number of shots in the training dataset in the Supplementary Material. To balance the effectiveness and training time, we choose 64 shots during the training phase.

3.4   Discussion

"Pneumonia" enables anomaly detection work. The reason that the class name "Pneumonia" enables anomaly detection work is based on two aspects. First, pneumonia is a prevalent and representative anomaly category in chest X-rays. Second, the inherent capacity of analogical generalization in the CLIP model ensures that the similarity between normal and pathological cases is less than the similarity between pneumonia and pathological cases. This aids in performing anomaly detection by measuring these similarity values. Besides, this analogical generalization capacity may extend the CLIP model to detect rare pathological categories.

Training efficiency. In terms of training parameters, the PPAD only has 0.046M learnable parameters, which is about 0.03% of the training parameters in the ViT-B/32 CLIP (151.32M). In terms of overall training time, the PPAD takes about 13 minutes in the 64-shot and 100-epoch setting. Overall, the proposed PPAD is efficient for practical deployment.

Limitation. Our anomaly synthesis method SAS appears to fall short in simulating specific diseases, such as cardiomegaly and fractures. This shortcoming becomes evident when considering its performance on the CheXpert [7] dataset,

which comprises 12 distinct diseases. The proposed method only achieves a modest improvement of 2.8% (average across three metrics) over the baseline method CheXzero [21] when tested on this dataset.

4 Conclusion

We propose PPAD, a position-guided prompt learning method for anomaly detection in chest X-rays. PPAD is based on CLIP-based methods to adapt the task-specific data to the pre-training data. Additionally, we introduce SAS for synthetic anomaly. Experiments on three datasets demonstrate that the PPAD achieves state-of-the-art performance. In the future, we intend to extend our PPAD method to achieve anomaly localization. Besides, we aim to refine our anomaly synthesis method SAS to incorporate a wider variety of anomalies.

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

References

1. Baugh, M., Tan, J., Müller, J.P., Dombrowski, M., Batten, J., Kainz, B.: Many tasks make light work: Learning to localise medical anomalies from multiple synthetic tasks. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 162–172 (2023)
2. 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 (2023)
3. 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)
4. Chen, X., He, Y., Xue, C., Ge, R., Li, S., Yang, G.: Knowledge boosting: Rethinking medical contrastive vision-language pre-training. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 405–415 (2023)
5. 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)
6. Gong, D., Liu, L., et al.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Int. Conf. Comput. Vis. pp. 1705–1714 (2019)
7. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proc. AAAI Conf. Artif. Intell. vol. 33, pp. 590–597 (2019)

8. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data **6**(1), 317 (2019)

9. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. cell **172**(5), 1122–1131 (2018)

10. Khattak, M.U., Rasheed, H., et al.: MaPLe: Multi-modal prompt learning. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 19113–19122 (2023)

11. 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)

12. 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 (2021)

13. Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T., Dinh, D.H., et al.: VinDr-CXR: An open dataset of chest x-rays with radiologist's annotations. Scientific Data **9**(1), 429 (2022)

14. Pellegrini, C., Keicher, M., Özsoy, E., Jiraskova, P., Braren, R., Navab, N.: Xplainer: From x-ray observations to explainable zero-shot diagnosis. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 420–429 (2023)

15. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: Proc. Int. Conf. Mach. Learn. pp. 8748–8763 (2021)

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. 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)

18. Sato, J., Suzuki, Y., Wataya, T., Nishigaki, D., Kita, K., Yamagata, K., Tomiyama, N., Kido, S.: Anatomy-aware self-supervised learning for anomaly detection in chest radiographs. iScience (2023)

19. 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)

20. 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)

21. Tiu, E., Talius, E., Patel, P., Langlotz, C.P., Ng, A.Y., Rajpurkar, P.: Expert-level detection of pathologies from unannotated chest x-ray images via self-supervised learning. Nature Biomedical Engineering **6**(12), 1399–1406 (2022)

22. 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)

23. You, K., Gu, J., Ham, J., Park, B., Kim, J., Hong, E.K., Baek, W., Roh, B.: CXR-CLIP: Toward large scale chest x-ray language-image pre-training. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention. pp. 101–111 (2023)

24. Zhang, S., He, L., Wang, D., Bao, H., Zheng, S., Liu, Y., Xiao, B., Li, J., Lu, D., Zheng, N.: Prosyno: context-free prompt learning for synonym discovery. Frontiers of Computer Science **19**(6), 196317 (2025)

25. 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)
26. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. Int. J. Comput. Vis. **130**(9), 2337–2348 (2022)
27. Zhou, K., Li, J., Luo, W., Li, Z., Yang, J., Fu, H., Cheng, J., Liu, J., Gao, S.: Proxy-bridged image reconstruction network for anomaly detection in medical images. IEEE Trans. Medical Imaging. **41**(3), 582–594 (2021)
28. Zhou, K., Xiao, Y., Yang, J., Cheng, J., Liu, W., Luo, W., Gu, Z., Liu, J., Gao, S.: Encoding structure-texture relation with p-net for anomaly detection in retinal images. In: Eur. Conf. Comput. Vis. pp. 360–377 (2020)