Anatomy-aware self-supervised learning for anomaly detection in chest radiographs

Junya Sato, Yuki Suzuki, Tomohiro Wataya, ..., Kazuki Yamagata, Noriyuki Tomiyama, Shoji Kido

kido@radiol.med.osaka-u.ac.jp

Highlights
Anomaly detection models in chest radiographs provide proper treatment decisions

Self-supervised learning is a key to enhance anomaly detection accuracy

We propose an anomaly detection model to exploit anatomical information

The results on three datasets showed that our method outperforms existing ones

Sato et al., iScience 26, 107086
July 21, 2023 © 2023 The Author(s).
hhttps://doi.org/10.1016/j.isci.2023.107086
Anatomy-aware self-supervised learning for anomaly detection in chest radiographs

Junya Sato,1,2,3,4 Yuki Suzuki,1 Tomohiro Wataya,1,2 Daiki Nishigaki,1,2 Kosuke Kita,1 Kazuki Yamagata,1,2 Noriyuki Tomiyama,2 and Shoji Kido1,*

SUMMARY
In this study, we present a self-supervised learning (SSL)-based model that enables anatomical structure-based unsupervised anomaly detection (UAD). The model employs an anatomy-aware pasting (AnatPaste) augmentation tool that uses a threshold-based lung segmentation pretext task to create anomalies in normal chest radiographs used for model pretraining. These anomalies are similar to real anomalies and help the model recognize them. We evaluate our model using three open-source chest radiograph datasets. Our model exhibits area under curves of 92.1%, 78.7%, and 81.9%, which are the highest among those of existing UAD models. To the best of our knowledge, this is the first SSL model to employ anatomical information from segmentation as a pretext task. The performance of AnatPaste shows that incorporating anatomical information into SSL can effectively improve accuracy.

INTRODUCTION
Deep learning has made significant progress in image recognition, including medical image recognition. A deep learning model learns informative image representations without manual feature extraction and pre-processing. It requires training with a large number of labeled images for accurate recognition.1–3 However, manual annotation is labor-intensive and time-consuming. Moreover, medical image annotation requires medical expert intervention and resources.4–6

Chest radiograph is the most frequently performed radiological examination in the world, with a reported 146 erect-view images acquired per 1,000 population annually.7 They are used to detect and classify various cardiac and pulmonary disorders. Numerous images containing disease-specific annotations are needed for disease classification. In addition, identifying diseases whose images are not used during model training is challenging. Likewise, prediction accuracy becomes low when there is little data available, as in the case of rare diseases.8 Hence, our goal is to solve the label deficiency and create a model with high diagnostic accuracy in chest radiographs.

An unsupervised anomaly detection (UAD) model is effective in such annotation and dataset size problems. This model learns only normal images and determines a characteristic threshold. The image exhibiting features beyond the threshold is classified as abnormal. We can easily collect normal images through medical examinations. Furthermore, UAD model can detect rare diseases unseen during model training. For these reasons, extensive studies have been conducted on UAD.9–19 Recently, models using deep learning for feature extraction and models using encoder-decoder networks such as variational autoencoder2,13 and generative adversarial network (GAN)14–19 have been actively employed.

Self-supervised learning (SSL)-based UAD enhances anomaly detection accuracy.10,19–22 SSL aims at learning data-specific semantic features by generating supervisory signals from a pool of unlabeled data without human annotation.23 In natural image classification datasets such as ImageNet, some SSL-based models have achieved accuracies comparable to conventional supervised learning.24–26 SSL-based models require an appropriate pretext task to acquire detailed characteristics of diseases (anomalies).27 Several SSL-based deep learning models have been proposed in the medical field.19,22,23,28–32 However, these models use the same SSL as natural images and disregard task-specific information in medical images.
For example, some popular SSL-based methods generate abnormal images by random pasting or masking. Classifying or repairing them helps the model learn image features. The generated abnormal areas have clear boundaries with the normal ones and occur where they should not appear anatomically. These augmentations are inadequate for models to learn the characteristics of actual anomalies. Thoracic diseases such as lung cancer and pneumonia often appear in the lung region. In addition, radiologists often detect disease by focusing on the relative position of the lungs to the other organs that make up their boundaries. When abnormal findings are present, the normal anatomical structures are often disrupted. Therefore, providing information about the lung region helps the model learn the representations of diseases.

In this study, we propose a novel SSL-based method, anatomy-aware pasting (AnatPaste) augmentation, which extracts the lung region and creates abnormal shadows within that region. AnatPaste employs normal images and generates real anomaly-like images by using organ location information. This feature assists deep learning models to recognize anomalies. Figure 1 depicts a comparison between AnatPaste and the existing SSL augmentation method. The existing method augments the images where anomalous areas are clearly demarcated and includes additional anomalous areas where the real anomalies do not occur. However, AnatPaste-developed anomalies appear real and are limited to the lung region.

AnatPaste does not require a specific deep learning model’s structure; therefore, it will help existing UAD models boost their accuracy. In addition, if proper segmentation is performed, AnatPaste can be extended to other imaging modalities such as CT and MRI and organs. AnatPaste has the potential to be widely applied to medical image analysis.

To evaluate our method for anomaly detection, we use a simple one-class classifier trained with AnatPaste. The UAD model is validated using three publicly available chest radiograph datasets. It exhibits significantly higher anomaly detection accuracy on all the datasets compared with existing UAD models. The contributions of our study are summarized as follows.

(1) AnatPaste is the first augmentation method that utilizes medical image segmentation as a pretext task.
AnatPaste can be applied to any model, downstream tasks, and even to other imaging modalities and organs by setting up segmentation methods.

The model performs the best on anomaly detection tasks on three chest radiograph datasets.

Related works

Anomaly detection

Anomaly detection, also known as outlier detection or novelty detection, involves the detection of data instances that deviate significantly from majority instances.34 Traditional models, such as one-Class support vector machine35 and support vector data description (SVDD),35 train a single normal data class and determine a threshold to detect the outliers. However, these models increase computational complexity when used with large and high-dimensional data. Recent studies have increasingly employed deep learning models for feature extractions. VGG trained on ImageNet1 is employed to obtain representations of X-ray security-screening images.36 Deep SVDD37 trains a representation to minimize the volume of a hypersphere of the normal samples.

Figure 2. Illustration of AnatPaste augmentation

A patch \(X_{\text{crop}}\) is randomly cropped from a normal chest radiograph \(X_{\text{normal}}\) and randomly pasted at another place \(X_{\text{paste}}\). Simultaneously, an anatomical mask \(X_{\text{mask}}\) is created using element-wise multiplication between the segmented lung region \(X_{\text{seg}}\) and the blurred region \(X_{\text{blur}}\). Furthermore, \(X_{\text{normal}}\) and \(X_{\text{paste}}\) are summed by using \(X_{\text{mask}}\). The anomalous region occurs only in the lung region in the augmented image.

Table 1. Average classification performances (AUC, accuracy, F1) and standard deviations yielded by baseline models and ours on Zhanglab, CheXpert, and RSNA dataset

|        | Zhanglab |                |                | CheXpert |                |                | RSNA  |
|--------|----------|----------------|----------------|----------|----------------|----------------|-------|
|        | AUC (%)  | ACC (%)        | F1 (%)         | AUC (%)  | ACC (%)        | F1 (%)         | AUC (%)| ACC (%)| F1 (%)|
| fAnoGAN| 62.4 ± 1.9 | 65.6 ± 1.9 | 74.4 ± 1.9 | 60.3 ± 0.7 | 57.1 ± 1.2 | 61.5 ± 3.3 | 60.1 ± 0.5 | 50.3 ± 0.2 | 66.7 ± 0.1 |
| Ganomaly| 75.1 ± 1.2 | 66.8 ± 5.7 | 78.5 ± 3.0 | 61.7 ± 4.6 | 56.5 ± 4.0 | 46.5 ± 20.6 | 65.7 ± 9.8 | 52.7 ± 2.1 | 67.4 ± 0.8 |
| CutPaste| 76.8 ± 2.9 | 75.2 ± 2.4 | 81.4 ± 1.6 | 58.0 ± 1.3 | 52.8 ± 2.0 | 64.5 ± 1.6 | 69.1 ± 2.0 | 51.6 ± 3.0 | 66.7 ± 0.1 |
| SALAD\(a\)| 82.7 ± 0.8 | 75.9 ± 0.9 | 82.1 ± 0.3 | 66.7 ± 0.8 | 66.7 ± 0.8 | 66.7 ± 0.8 | 66.7 ± 0.8 | 66.7 ± 0.8 | 66.7 ± 0.8 |
| SQUID  | 85.9 ± 2.5 | 82.4 ± 1.4 | 74.6 ± 2.7 | 66.7 ± 9.8 | 65.5 ± 3.4 | 60.3 ± 7.9 | 60.5 ± 2.5 | 66.4 ± 1.3 | 54.0 ± 2.0 |
| AnatMix| 91.4 ± 1.3 | 83.0 ± 1.9 | 86.8 ± 1.1 | 76.0 ± 1.8 | 65.4 ± 3.8 | 71.2 ± 1.4 | 80.4 ± 1.4 | 72.3 ± 1.8 | 75.1 ± 0.7 |
| Ensemble| 92.1      | 83.2          | 86.4          | 78.7     | 69.2          | 73.5          | 81.9   | 72.8   | 75.6   |

Ensemble means the average anomaly scores of five AnatPaste models.

\(a\)The results are taken from the original paper.
Several models using encoder-decoder networks such as autoencoder have also been proposed to learn normal data distributions. AnoGAN16 is the first generative adversarial network model for anomaly detection. AnoGAN identifies a query image as abnormal when there is a large difference between the query image and the image generated by learning only normal images. fAnoGAN15 is a faster and more accurate version of AnoGAN owing to the introduction of an encoder. Ganomaly38 and SALAD19 improve anomaly detection accuracy using a loss function with reconstructed images and latent spatial variables. These models are effective on some datasets, but reconstruction errors make it challenging to capture minor anomalies. 39,40 SQUID18 employs space-aware memory queues, as medical images are structured owing to consistent imaging protocols. However, SQUID requires simultaneous training of its components (one discriminator, encoder, and two generators). Therefore, it encounters structural complexity.

**Self-supervised learning**

SSL is an approach that learns semantically useful features for a task through a pretext task without manual annotations. SSL generates a supervisory signal from a pool of unlabeled data for the downstream task. The key challenge for SSL is setting up an appropriate task that allows the model to learn the features of the dataset well. For natural images, various SSL methods have been proposed, such as contrastive learning,24,25 rotation prediction,41 jigsaw puzzle,42 denoising autoencoder,43 and in-painting.33 Although these methods have been widely used in medical imaging,19,22,23,28–32 they just adopt the SSL method proposed for natural images without any modification. As an SSL method specialized for medical images, Zhang et al.44 treat 3D images such as CT and MRI as a set of 2D slice images and perform a pretext task of predicting the spatial position of the 2D slices. Tian et al.10 devise MedMix, a pretext task based on contrastive learning designed for medical images. However, they do not consider the anatomical structure of organs. CutPaste21 is an SSL method specialized for anomaly detection. It generates an anomaly image by cutting a part of a normal image and pasting it to another location in the same image. It has achieved state-of-the-art performances in anomaly detection for the industrial product dataset MVTe-cAD.45 However, many medical images have complex structures and often need medical expertise, making it challenging to determine anomalies. To achieve accurate anomaly detection in medical images, the method should be redesigned such that the model can understand the structure of complex anatomy.

**Self-supervised learning in chest radiographs**

Many SSL methods have been proposed specifically for chest radiographs. Most of them are based on contrastive learning.29,46–48 Typically, contrastive learning focuses on learning global information about an image through augmentation, such as rotation and noise, resulting in an insufficient representation of local information. Zhou et al.49 proposed an image restoration task that preserves and exploits pixel-level local information.

Other methods have also been proposed for SSL using metadata associated with images.37,48,50,51 Vu et al. improved contrastive learning performance by using images of the same patient from the metadata as a positive pair.47 In addition, disease-based representations can be learned by using free-text radiology reports. Zhang et al.48 proposed a method that uses images and paired reports as input for contrastive learning to obtain medical visual representations without requiring additional expert input. Huang

Figure 3. Receiver operating characteristic curve (ROC) of the comparison methods

AnatPaste and other anomaly detection methods are evaluated using three publicly available chest radiograph datasets.
et al. proposed a framework that contrasts image subregions with paired report words to learn local and global representations of images. Zhou et al. collected a series of images and paired texts of the same patient and fused their features. However, few previous SSLs have focused on organ structures in the image. Many anomalies, such as atelectasis, pulmonary emphysema, and consolidation, appear within the lungs. Therefore, providing anatomical information about the lung region allows efficient learning of image features related to the disease.

RESULTS
AnatPaste overview
AnatPaste is an augmentation method that enables a model to learn the anatomical structural information. Figure 2 depicts an AnatPaste augmentation method. Our proposed model first uses AnatPaste to train a convolutional neural network model. Further, the trained model works as a feature extractor to detect anomalies. AnatPaste, the SSL training setup, and the anomaly score calculations are in the STAR Methods section.

Anomaly detection performance compared with the existing methods
The accuracy of AnatPaste was benchmarked across three datasets: Zhanglab, CheXpert, and RSNA dataset. The performance was compared against four traditional anomaly detection models: fAnoGAN, Ganomaly, CutPaste, and SQUID. Details regarding these datasets and conventional methods can be found in the STAR Methods section. Performance was validated using the receiver operating characteristic (ROC) curve, area under the curve (AUC), accuracy, and F1-score metrics. Table 1 and Figure 3 compare the contemporary methods’ AUC, accuracy, F1-scores, and ROC curves on the three datasets. Our method exhibits the best performances of all the contemporary methods. On the Zhanglab dataset, AnatPaste achieves 14.6% higher AUC than the conventional CutPaste (76.8%). SQUID achieves the best results among existing methods, with an AUC of 85.9% on the Zhanglab dataset. AnatPaste has 5.5% higher AUC than SQUID and achieves the best accuracy and F1-score. Further, AnatPaste exhibits 18.0% and 11.3% higher AUC than the CutPaste algorithm on the CheXpert dataset and RSNA dataset, respectively, indicating that our method exhibits higher accuracy than the conventional models. Furthermore, the ensemble of the five AnatPaste models (Table 1) achieves an accuracy improvement of 0.7%, 2.7%, and 1.5% on Zhanglab, CheXpert, and RSNA, respectively.

Anomaly score prediction
Anomaly scores are calculated for each image to assess their tendencies in normal and abnormal labels. The output of the model with the highest AUC score is shown in Figure 4. The threshold is set to 0.067.
based on the best F1-score. The images in green boxes are correctly predicted by the model, and those in red boxes are not. The more the areas of abnormal white shadows within the lung field, the higher the anomaly score tends to be. By contrast, the model often incorrectly predicts normal labeled images with pacemakers and other artifacts.

**Visualization**

Figure 5 depicts the Grad-CAM visualization of the annotated areas in two images. The last convolutional layer’s feature map of the pretext model is used to calculate the gradient. The original images (left panels) are enhanced using the Grad-CAM (right panels). The areas judged to be abnormal are highlighted in red. In the abnormal case, the red-colored highlighted areas appear corresponding to consolidation in the lung. By contrast, in the normal case, there are no red-colored annotated areas, and the model does not indicate a specific area.

**Table 2. Anomaly detection performance by disease in CheXpert dataset**

| Disease          | AUC(%) | ACC(%) | F1(%) |
|------------------|--------|--------|-------|
| Atelectasis      | 71.2 ± 1.8 | 63.1 ± 3.2 | 68.7 ± 0.5 |
| Cardiomegaly     | 73.3 ± 1.8 | 62.9 ± 3.9 | 68.4 ± 1.6 |
| Consolidation    | 78.2 ± 2.1 | 65.6 ± 4.1 | 71.5 ± 2.0 |
| Edema            | 76.5 ± 1.7 | 64.7 ± 4.3 | 70.5 ± 2.0 |
| Pleural effusion | 80.0 ± 2.4 | 66.8 ± 4.4 | 72.8 ± 2.0 |
### Evaluation by disease

Table 2 shows the results of AnatPaste by disease on the CheXpert dataset. 300 images per disease and 300 normal images are randomly selected to validate performances. Pleural effusion shows the best abnormality detection accuracy, followed by consolidation and edema. The AUC of consolidation, edema, and pleural effusion achieved via supervised learning in the original paper is similar by approximately 93%. Compared to these results, AnatPaste has a relatively high accuracy in detecting pleural effusion.

### Ablation study

We perform an ablation study to validate the essentiality of the components of AnatPaste, which are segmented lung region \( X_{\text{seg}} \) and Gaussian blur \( X_{\text{blur}} \), as shown in Figure 2. Figure 6 shows the details of each condition in the ablation study. In “w/o anatomy awareness,” all \( X_{\text{seg}} \) pixel values are set to 1, resulting in an anatomical mask without information about the lung region. In the absence of Gaussian blur, setting all \( X_{\text{blur}} \) pixel values to 1 produces an image in which the edges of the patch to be pasted remain unblurred. The CheXpert dataset is used for the validation. The evaluation metrics (AUC, accuracy, and F1-score) are calculated. In Table 3, the AUCs without segmentation and without Gaussian blur are 70.7% and 60.2%, respectively, which are 5.3% and 15.8% lower than that of AnatPaste, respectively, suggesting that both the components improved accuracy.

### DISCUSSION

Chest radiography is the most common and minimally invasive screening test in medical diagnosis. There are various chest diseases to be detected, but the shortage of labeled datasets has prevented accurate training of disease detection models. The UAD model not only solves such problems but also serves as a valuable computer-aided diagnosis system that provides appropriate treatment decisions.

In this study, we propose a simple and effective SSL-based method that detects chest radiograph anomalies, using anatomical structures in the lung fields. Although some supervised learning methods segment lung...
fields and classify them, few SSL methods provide organ location information. Our method takes advantage of the medical knowledge that most abnormalities appear in the lung field region and augments only the segmented lung fields. The images generated by our method are more similar to the actual anomalies than those generated by existing methods. Our model trained with AnatPaste exhibits high feature representation capability even without disease labels and significantly improves anomaly detection accuracy at least by 5.5%, 14.3%, and 11.3% in terms of AUC compared with existing methods on three publicly available datasets. Our results indicate that incorporating disease and anomaly-specific medical expertise in the training can enable accurate medical image analysis.

AnatPaste can be applied as a pretext task regardless of the structure of the deep learning model; therefore, further performance improvement can be expected by incorporating it into existing anomaly detection models. In addition, AnatPaste’s feature representation capability can be used not only for anomaly detection but also for downstream tasks such as disease classification and segmentation through supervised learning. Thus, AnatPaste can be applied to a variety of medical image analysis and has the potential to contribute to improving accuracy.

Limitations of the study
This study has a limitation regarding segmentation methods. We do not adopt deep-learning-based lung segmentation methods because they require annotations and are outside the scope of the unsupervised

| Ablation study comparing the full AnatPaste, AnatPaste without image segmentation, and AnatPaste without Gaussian blur |
|--------------------------------------------------|
|                      | AUC (%) | ACC (%) | F1 (%) |
|-----------------------|---------|---------|--------|
| Full AnatPaste        | 76.0 ± 1.8 | 65.4 ± 3.8 | 71.2 ± 1.4 |
| w/o Segmentation      | 70.7 ± 1.6 | 55.7 ± 5.9 | 66.2 ± 0.9 |
| w/o Gaussian blur     | 60.2 ± 2.9 | 52.5 ± 2.7 | 65.1 ± 2.5 |

Algorithm 1. Lung segmentation

Input: Normal chest radiograph
Initialize t of a threshold to be considered as lung region

Output: Binarized lung segmentation image

step:1 Equalize local image histogram using CLAHE.

step:2 Binarize histogram equalized image using Otsu’s thresholding.

step:3 Apply morphological opening to remove noise.

step:4 Remove objects connected to the image border.

step:5 Apply morphological dilation to cover whole lung area: Xseg

step:6 Compute sets of data points (R1 ... RK) using connected-components labeling. Each set has same label.

step:7 for k = 1 : K
  if |Rk| < t :
    Xseg = g(Xseg, Rk, 0)
  end if
end for
where g(X, R, p) is a function that sets pixel value p for the pixels of the input image X at coordinates in the R and returns updated image.

step:8 Return Xseg
learning framework. Since AnatPaste is applied only to normal images, we think it is possible to extract lung regions in all images. If an accurate segmentation method is developed, further performance improvements are expected. In addition, our segmentation method is limited to lung fields in chest radiographs. AnatPaste can be extended to other organs and modalities by employing appropriate segmentation methods.

**Conclusions**

We proposed a novel SSL-based method for chest radiograph anomaly detection. Our objective was to adopt carefully designed augmentations such that abnormal areas are limited to a single organ, thus allowing the model to recognize anomalies in the context of the anatomical structure. Experimental results on three chest radiograph benchmark datasets demonstrated that our method outperformed existing unsupervised anomaly detection methods. AnatPaste can be applied to a variety of other downstream tasks in chest radiographs as a pretext task, regardless of the model type. Furthermore, the proposed method can be applied to different organs and modality by setting up an appropriate segmentation. Our method can potentially improve the accuracy of deep learning models in medical imaging.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - AnatPaste augmentation
  - Anomaly score
  - Image sources and details
  - Experimental settings
  - Baseline & evaluation

**ACKNOWLEDGMENTS**

This work was supported by JSPS KAKENHI Grant Number 21H03840.

**AUTHOR CONTRIBUTIONS**

Conceptualization, J.S., N.T., and S.K.; Methodology, J.S., Y.S., T.W., and K.K.; Formal Analysis, J.S., Y.S., and D.N.; Writing – Original Draft, J.S.; Writing–Review & Editing, J.S., Y.S., T.W., K.K., and K.Y.; Funding Acquisition, N.T. and S.K.; Resources, J.S.; Supervision, N.T. and S.K.

**DECLARATION OF INTERESTS**

Authors declare that they have no competing interests.
from non-small cell lung cancer histopathology images using deep learning. Nat. Med. 24, 1559–1567. https://doi.org/10.1038/s41591-018-0177-5.

3. Langlotz, C.P., Allen, B., Erickson, B.J., Kalpathy-Cramer, J., Bigelow, K., Cook, T.S., Flanders, A.E., Lungren, M.P., Mendelson, D.S., Rudie, J.D., et al. (2019). A roadmap for foundational research on artificial intelligence in medical imaging. From the 2018 NIH/RSNA/AACR/The Academy Workshop. Radiology 291, 781–791. https://doi.org/10.1148/radiol.2019190613.

5. You, C., Zhou, Y., Zhao, R., Staib, L., and Duncan, J.S. (2022). SimCvD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Med. Imaging 41, 2228–2237. https://doi.org/10.1109/TMI.2022.3161829.

6. Xu, Y., Zhu, J.-Y., Chang, E.I.-C., Lai, M., and Tu, Z (2014). Weakly supervised histopathology image segmentation and classification. Med. Image Anal. 18, 591–604. https://doi.org/10.1016/j.media.2014.01.010.

7. United Nations Scientific Committee on the Effects of Atomic Radiation (2010). Sources and Effects of Ionizing Radiation, United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCAR) 2008 Report, Volume I: Report to the General Assembly, with Scientific Annexes A and B - Sources (UN). https://www.unscear.org/docs/reports/2008/09-86753_Report_2008_GA_Report.pdf.

8. Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. (2015). How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?. Preprint at arXiv. https://doi.org/10.48550/arXiv.1511.06348.

9. Chen, Y., Tian, Y., Pang, G., and Carneiro, G. (2021). Deep one-class classification via interpolated Gaussian descriptor. In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 383–392. https://doi.org/10.1609/aaai.v35i03.19915.

10. Tian, Y., Liu, F., Pang, G., Chen, Y., Liu, Y., Verjans, J.W., Singh, R., and Carneiro, G. (2021). Self-supervised multi-class pre-training for unsupervised anomaly detection and segmentation in medical images. Preprint at arXiv. https://doi.org/10.48550/arXiv.2109.01303.

11. Zhang, J., Xie, Y., Pang, G., Liao, Z., Verjans, J., Li, W., Sun, Z., He, J., Li, Y., Shen, C., and Xia, Y. (2021). Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection. IEEE Trans. Med. Imaging 40, 879–890. https://doi.org/10.1109/TMI.2020.3040950.

12. Baur, C., Denner, S., Wiestler, B., Navab, N., and Albarqouni, S. (2021). Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. Med. Image Anal. 69, 1019522. https://doi.org/10.1016/j.media.2020.1019522.
Aphinyanaphongs, Y., Yakoubova, N., and Moore, W. (2021). COVID-19 prognosis via self-supervised representation learning and multi-image prediction. Preprint at arXiv. https://doi.org/10.48550/arXiv.2101.04909.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A.A. (2016). Context encoders: feature learning by inpainting. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2536–2544. https://doi.org/10.1109/CVPR.2016.278.

Pang, G., Shen, C., Cao, L., and Hengel, A.V.D. (2022). Deep learning for anomaly detection: A review. ACM Comput. Surv. 54, 1–38. https://doi.org/10.1145/3439950.

Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R.C. (2001). Estimating the support of a high-dimensional distribution. Neural Comput. 13, 1443–1471. https://doi.org/10.1162/089976601750246965.

Andrews, J., Tanay, T., Morton, E.J., and Gidaris, S., Singh, P., and Komodakis, N. (2016). Unsupervised representation learning of fine-grained body part recognition. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 578–582. https://doi.org/10.1109/ISBI.2017.7950587.

Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. (2019). MVtec AD-A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9592–9600. https://doi.org/10.1109/CVPR.2019.00982.

Zhou, H.-Y., Lu, C., Yang, S., and Yu, Y. (2021). Preservational learning improves self-supervised medical image models by reconstructing diverse contexts. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3499–3509. https://doi.org/10.1109/CVPR48922.2021.00348.

Yu, Y.N.T., Wang, R., Balachandar, N., Liu, C., Ng, A.Y., and Rajpurkar, P. (2021). MedAug: Contrastive learning leveraging patient metadata improves representations for chest X-ray interpretation. In Proceedings of the 6th Machine Learning for Healthcare Conference Proceedings of Machine Learning Research, K. Jung, S. Yeung, M. Sendak, M. Snojed, and R. Ranganath, eds. (MLR), pp. 755–769.

Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., and Langlotz, C.P. (2022). Contrastive learning of medical visual representations from paired images and text. In Proceedings of the 7th Machine Learning for Healthcare Conference, 182 Proceedings of the 7th Machine Learning for Healthcare Conference, pp. 2–25.

Zhou, H.-Y., Lu, C., Chen, C., Yang, S., and Yu, Y. (2023). A unified visual information preservation framework for self-supervised pre-training in medical image analysis. IEEE Trans. Pattern Anal. Mach. Intell. 45, 8020–8035. https://doi.org/10.1109/TPAMI.2023.3234002.

Zhou, H.-Y., Chen, X., Zhang, Y., Luo, R., Wang, L., and Yu, Y. (2022). Generalized radiomics representation learning via cross-supervision between images and free-text radiology reports. Nat. Mach. Intell. 4, 32–40. https://doi.org/10.1038/s42256-021-00245-9.

Huang, S.-C., Shen, L., Lungen, M.P., and Yeung, S. (2021). GlORiA: A multimodal global-local representation learning framework for label-efficient medical image recognition. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3922–3931. https://doi.org/10.1109/ICCV48922.2021.00391.

Kermayer, D.O.S., Goldbaum, M., Cai, W., Valentin, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172, 1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.

The DeepRadiology Team (2018). Pneumonia detection in chest radiographs. Preprint at arXiv. https://doi.org/10.48550/arXiv.1811.08939.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization (2017 IEEE International Conference on Computer Vision (ICCV)), pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.

Kumarasinghe, H., Kolonne, S., Fernando, C., and Meedeniya, D. (2022). U-Net based chest X-ray segmentation with ensemble classification for Covid-19 and pneumonia. Int. J. Onl. Eng. 18, 161–175. https://doi.org/10.1016/j.iopje.2020.03.006.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Graphics Gems, P.S. Heckbert, ed. (Academic Press), pp. 474–485. https://doi.org/10.1016/S0734-189X(87)80087-0.80087-X.

Otsu, N. (1978). A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66. https://doi.org/10.1109/TSMC.1979.4301676.

Jamil, N., Sembok, T.M.T., and Bakar, Z.A. (2018). Unsupervised representation learning by predicting image rotations. Preprint at arXiv. https://doi.org/10.48550/arXiv.1803.0728.

Noroozi, M., and Favaro, P. (2016). Unsupervised learning of visual representations by solving jigsaw puzzles. In Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9910, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. (Springer), pp. 69–84. https://doi.org/10.1007/978-3-319-46466-4_5.
common thorax diseases. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE), pp. 3462–3471. https://doi.org/10.1109/cvpr.2017.369.

63. He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. Preprint at arXiv. https://doi.org/10.48550/arXiv.1512.03385.

64. Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.

65. Loshchilov, I., and Hutter, F. (2016). SGDR: Stochastic gradient descent with warm restarts. Preprint at arXiv. https://doi.org/10.48550/arXiv.1608.03983.

66. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In Proceedings of Advances in Neural Information Processing Systems 32 (NeurIPS 2019, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, eds. (Neural Information Processing Systems Foundation).
STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited data      |        |            |
| Zhanglab dataset    | Kermany et al.52 | https://data.mendeley.com/datasets/rscbjbr9s/2 |
| CheXpert dataset    | Irvin et al.2 | https://stanfordmlgroup.github.io/competitions/chexpert/ |
| RSNA dataset        | RSNA Pneumonia Detection Challenge. Radiological Society of North America. 2018 | https://www.kaggle.com/competitions/rsna-pneumonia-detection-challenge/data |
| Software and algorithms |        |            |
| python 3.9.7        | Python Software Foundation | https://www.python.org/ |
| pytorch 1.10.1      | Pytorch software | https://pytorch.org/ |
| fAnoGAN             | Schlegl et al.15 | https://github.com/dbbbbm/f-AnoGAN-PyTorch |
| Ganomaly            | Akcay et al.38 | https://github.com/samet-akcay/ganomaly |
| CutPaste            | Li et al.31 | https://github.com/Runinho/pytorch-cutpaste |
| SQUID               | Xiang et al.18 | https://github.com/tiangexiang/SQUID |
| Other               | Github | https://github.com/jun-sato/AnatPaste |

RESOURCE AVAILABILITY

Lead contact
Further information and requests for model implementation should be directed to and will be fulfilled by the lead contact, Junya Sato (j-sato@radiol.med.osaka-u.ac.jp).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- This paper analyses existing, publicly available data. Detailed descriptions are listed in the key resources table.
- Source code and usage for implementing AnatPaste are publicly available online. The URL is listed in key resources table.

METHOD DETAILS

AnatPaste augmentation
First, the following lung segmentation procedure is performed on the normal chest radiographs. CLAHE56,57 is used for image normalization, which improves visibility by homogenizing the image histograms and enhancing the local region contrast. The image is then binarized using Otsu’s method.58 Furthermore, noise is removed using morphological transformation.59 The areas bordering the edges of the image are removed as background. The areas above a certain size are considered lung fields (Xseg in Figure 2).

Next, a patch (Xcrop in Figure 2) is created by randomly cropping the original image Xnormal. The size of Xcrop is the same as the original one, and the region other than the cropped region is valued at zero. The patch is randomly pasted at another place (Xpaste). Then, an oval or rectangle is randomly drawn in the same position as the pasted position in Xpaste (Xblur). The figure’s edge is blurred by a Gaussian blur to mimic the actual anomaly. This position is determined based on the following criteria: a) part of the figure should
be included in the lung region, and b) the figure size should be smaller than the patch. An anatomical mask is created by calculating the element-wise multiplication between the segmented lung region and the blurred region ($X_{\text{mask}}$). Furthermore, $X_{\text{normal}}$ and $X_{\text{paste}}$ are summed by using an element-wise sum with $X_{\text{mask}}$. The pixel values of the blurred region and the Gaussian blurred radius are sampled randomly from uniform distributions of $[0.59, 1]$ and $[0, 1.5]$, respectively.

Given a normal image ($X_{\text{normal}}$), a patch ($X_{\text{patch}}$), and an anatomical mask ($X_{\text{mask}}$), the generated anomalous image ($X_{\text{anomaly}}$) is expressed as:

$$X_{\text{anomaly}} = X_{\text{normal}} \ast (1 - X_{\text{mask}}) + X_{\text{patch}} \ast X_{\text{mask}}$$  \hspace{1cm} (Equation 1)

The loss function is expressed as:

$$L = \frac{1}{N} \sum_{i=1}^{N} \{ \text{Lce}(f(X_i), 0) + \text{Lce}(f(Aug(X_i)), 1) \}$$  \hspace{1cm} (Equation 2)

wherein the ith of $N$ normal image data is input to $f$, a CNN model that makes binary predictions. Aug denotes AnatPaste augmentation. The original image is labeled 0, and the augmented one is labeled 1. Lce denotes the cross entropy loss. Details of the lung segmentation is presented in Algorithm 1.

### Anomaly score

We use kernel density estimator (KDE) to calculate anomaly scores as reported by Li et al. Given the set of feature representations that the CNN outputs, KDE is computed as:

$$A(z) = \frac{1}{N} \sum_{i=1}^{N} K(z - z_i)$$  \hspace{1cm} (Equation 3)

where

$$K(z) = \exp \left( -\frac{\|z\|^2}{2} \right)$$  \hspace{1cm} (Equation 4)

The output anomaly scores are normalized from zero to one. A threshold is set to maximize the F1-score for each model.

### Image sources and details

Three publicly available chest radiograph datasets are used to verify performances. All the images are resized to 256 $\times$ 256 pixels.

#### Zhanglab dataset

This dataset includes 6480 frontal view chest radiographs. Each image is labeled as normal or pneumonia (an anomalous condition). This dataset is officially split into training (1349 normal and 3883 pneumonia images) and test (234 normal and 390 pneumonia images) sets in advance. We choose 200 images (100 from normal and 100 from pneumonia) randomly from the training data for validating model selection and setting thresholds.

#### CheXpert dataset

This dataset consists of 224,316 chest radiographs from 65,240 patients. The radiographs are annotated in 14 pathological categories, using radiology diagnostic reports. A total of 5,466 normal posterior-anterior (PA) images are used for training. We focus on CheXpert competition task diseases (Pleural Effusion, Cardiomegaly, Consolidation, Edema, and Atelectasis). A total of 300 normal and 300 abnormal images are selected randomly from the training dataset for the test. We select at least 50 images per disease. In the officially split validation dataset, 14 normal and 19 abnormal PA images are used for validation as suggested by Xiang et al. We ensure that patients do not overlap. We use this dataset in the ablation study. Some of the images are used in visualization.

#### RSNA dataset

This is the Radiological Society of North America (RSNA) Pneumonia Detection Challenge dataset, which is a subset of the NIH chest radiograph dataset. It consists of 9,790 normal and 19,894 abnormal (labeled
as no lung opacity/not normal or lung opacity) chest radiographs. Of these, a total of 6,614 normal PA images are used for training, 300 normal and 300 abnormal images for validation, and 300 normal and 300 abnormal images for testing.

**Experimental settings**

We use a one-class classifier with ResNet18 for feature extraction. We add three fully connected layers to this backbone as the projection head. The batch size is set as 64. The network is optimized using an SGD optimizer, with an initial learning rate, momentum, and weight decay of 0.03, 0.9, and 0.00003, respectively. The learning rate is scheduled using a cycle cosine annealing schedule. The model is trained for 256 epochs, and the weights are saved when the AUC for the validation dataset is the highest. The weights are loaded to evaluate performances on the test dataset. The entire framework is built on PyTorch with an NVIDIA RTX 3090 24 GB GPU.

**Baseline & evaluation**

Several unsupervised anomaly detection models have been proposed. We compare the following four contemporary models.

1. fAnoGAN is the state-of-the-art UAD model in for medical images. Its three-component architecture contains an encoder, decoder, and discriminator. fAnoGAN is trained using only normal images. Abnormal images are detected when the decoder fails to restore them, leading to increased reconstruction errors.

2. Ganomaly is another GAN-based UAD model. It employs three losses (adversarial loss, contextual loss, and encoder loss) in anomaly detection. It has exhibited accurate anomaly detection in several datasets, including the X-ray security-screening datasets.

3. CutPaste is a one-class classifier model with SSL proposed for anomaly detection in industrial products. The model employs a pretext task to extract a region from a normal image, paste it on another part of the image, and classify the image as an anomaly. We use the CutPaste-Scar variant, which exhibits the best accuracy in the original study.

4. SQUID uses a transformer based in-painting module that preserves spatial information in the image. It has achieved high anomaly detection performance on the Zhanglab and CheXpert datasets.

We employ model-specific learning parameters as reported in the respective studies. For a fair comparison, all methods except augmentation are identical in our model and the CutPaste algorithm. SALAD is validated on the Zhanglab dataset. As the original code is unavailable, we compare it based on the accuracy reported in the respective paper. We present the average scores out of five repeated runs. The validation threshold is determined based on the best F1-score.