Region and Global-Specific PatchCore based Anomaly Detection from Chest X-ray Images

Hyunbin Kim[1], and Junchul Chun[2*]
[1] ADDD
Yeoksam-Ro 175, Kang Nam, Seoul, Korea
[e-mail: nu_start@naver.com]
[2] Department of Computer Science, Kyoggi University
Suwon, Korea
[e-mail: jcchun@kgu.ac.kr]
[*]Corresponding author: Junchul Chun

Abstract

This paper introduces a method aimed at diagnosing the presence or absence of lesions by detecting anomalies in Chest X-ray images. The proposed approach is based on the PatchCore anomaly detection method, which extracts a feature vector containing location information of an image patch from normal image data and calculates the anomaly distance from the normal vector. However, applying PatchCore directly to medical image processing presents challenges due to the possibility of diseases occurring only in specific organs and the presence of image noise unrelated to lesions. In this study, we present an image alignment method that utilizes affine transformation parameter prediction to standardize already captured X-ray images into a specific composition. Additionally, we introduce a region-specific abnormality detection method that requires affine-transformed chest X-ray images. Furthermore, we propose a method to enhance application efficiency and performance through feature map hard masking. The experimental results demonstrate that our proposed approach achieved a maximum AUROC (Area Under the Receiver Operating Characteristic) of 0.774. Compared to a previous study conducted on the same dataset, our method shows a 6.9% higher performance and improved accuracy.

1. Introduction

$\mathbf{V}$ision-based image analysis techniques, including classification, localization, object detection, and semantic segmentation, have found widespread use in X-ray images, especially in PA (Posterior-Anterior) X-ray images for diagnosing lesions. Recent research has presented several cases applying deep neural network-based supervised learning methods to large-scale medical image datasets [1, 2, 3]. Notably, there is a diagnostic method in the attention-based series that utilizes Contrastive Learning [4]. J. Lui performs pre-processing on chest images to generate patch-wise attention. Additionally, self-supervised learning (SSL) approaches using pseudo-labels have been proposed for diagnosis through self-evolution [5].

Anomaly detection typically involves identifying data or data points that deviate from the normal data category. Unlike conventional supervised learning-based AI classifiers that focus on specific target objects, anomaly detectors identify objects that deviate from the defined normal data range. In the context of medical images, anomaly detection pertains to the task of diagnosing the presence or absence of lesions within images. Thus, the baseline of the anomaly detection method was designed to detect anomalies from the entire area of the input image. However, to enhance the accuracy of the anomaly detection method, it is crucial to consider specific locations where anomalies can occur. Therefore, in this paper, we propose a novel approach to detect anomalies in both local and global areas of the X-ray image at the same time. In our previous work [6], we introduced an anomaly detection method that extracted patch features from a pre-trained model, using the histogram equalization technique known as CLAHE (Contrast Limited Adaptive Histogram Equalization).

Chest X-ray images capture multiple organs simultaneously, leading to the occurrence of various lesions concurrently in these images. In this work, we present a comprehensive approach to detect both region-specific lesions, occurring in specific organs, and global lesions that may manifest in various areas. The detection of region-specific lesions involves applying anomaly detection to individual organs separately. To achieve this, we extract localized patch features from aligned chest X-ray images and apply anomaly detection to each region. Additionally, we conduct anomaly detection for global lesions and then combine the two results to comprehensively detect anomalies.

Our proposed method is based on the PatchCore anomaly detection approach, which entails extracting feature vectors containing spatial information from normal image data. We then calculate distances from the normal vectors to identify anomalies. However, PatchCore's comprehensive definition of normal features across the entire image area poses challenges in medical image processing, where diseases may occur only in specific organs and unrelated image noise can be present. Therefore, we propose a region-specific anomaly detection method, which is capable of diagnosing lesions occurring solely in specific regions while ignoring noise. To accomplish this, we introduce a region-specific anomaly detection method that aligns X-ray images acquired under different views through the prediction of affine transformation parameters. We then utilize affine-transformed chest X-ray images. In this process, we present a method to enhance the efficiency and performance of the system using feature map hard masking.

2. Related Works

Anomaly detection is the task of identifying abnormal data or data points that deviate from the normal data distribution defined by the model. Abnormal elements refer to those belonging to the complement set of the defined normal set. This type of detection is useful when

conventional classification systems encounter challenges, such as difficulty in obtaining abnormal data samples or defining features for abnormal data. Anomaly detection is performed under such circumstances, and as a result, the training dataset typically consists only of normal samples, leading to the predominant use of semi-supervised learning-based methods. In the context of digital image data, anomaly detection can be broadly divided into two tasks. Firstly, binary classification inference can be performed at the image level. Additionally, anomaly estimation can be conducted at the pixel or patch level, resulting in heat map-shaped outcomes. These heat map results can then be transformed into binary masks through thresholding, allowing for the segmentation of areas estimated as anomalies.

Anomaly detection methods related to image reconstruction, classification, feature matching, and probabilistic models are broadly categorized, as illustrated in **Fig. 1**.
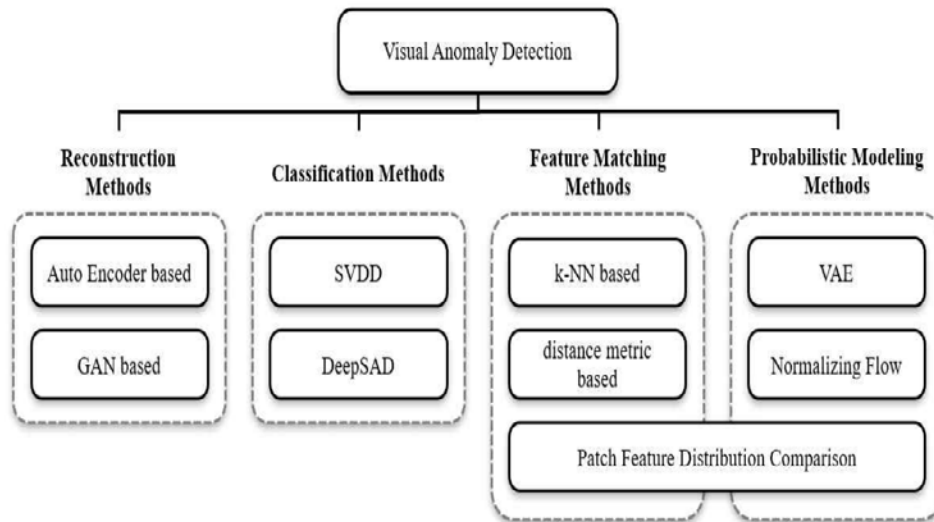

Fig. 1. The classification of anomaly detection method

The image reconstruction method involves restoring all input images to the shape of learned normal images. During the inference process, a normal image is output as it is, while an abnormal image results in an output with abnormal characteristics removed. The Image Difference method is then employed to compare the generated image with the input image, identifying regions with differences as predicted anomalies. A prominent research case in image anomaly detection using the reconstruction method is the use of auto-encoders [7]. In this approach, the encoder extracts a latent vector, and the decoder reconstructs an image with normal features. The method assumes that the multivariate distribution of the latent vector follows a Gaussian distribution and extends the research by incorporating the Variational Auto Encoder (VAE), which adjusts the distribution to improve performance. VAE can be seen as a fusion of image reconstruction techniques and probabilistic model-based methods. CAE (Contractive Auto Encoder) [8] addresses the zero-derivation problem that occurs in conventional auto-encoders. On the other hand, AnoGAN (Anomaly Detection using Generative Adversarial Network) [9], primarily used for image generation, has been adapted for retinal image anomaly detection by incorporating Residual Loss into the GAN structure.

The classification method is based on the Support Vector Machine (SVM) approach, where the feature vectors extracted from normal images are adjusted to cluster as closely as possible around a single point. A well-known research case is DeepSVDD (Deep Support Vector Data Description) [10]. The feature comparison method involves extracting features using a pre-trained deep neural network model and directly comparing them with the input data during the inference stage [11]. The feature extraction primarily relies on the Wide-ResNet-50 model, known for its simple block structure inherited from the ResNet family, making it suitable for extracting mid-level features and localized information from patches [12,13]. However, since the extracted features are represented as high-dimensional vectors or feature maps, comparing features incurs significant computational costs. To mitigate this, applying nearest neighbor search algorithms can dramatically reduce the number of comparison targets [14].

The proposed method utilizes feature comparison-based anomaly detection techniques to specifically detect lesions that primarily appear in certain regions.

3. The Proposed Anomaly Detection from X-ray Images

The overall procedure for detecting anomalies from X-ray images is depicted in **Fig. 2**. The inference data processing flow of the medical image anomaly detection system proposed in this paper is carried out in the following order:

- Alignment of chest X-ray images.
- Extraction of feature vectors containing positional information from the aligned images.
- Extraction of feature vectors containing positional information from the aligned images.
- Measurement of the average distance from the selected normal feature vectors, which is used to determine the anomaly score.

Fig. 2 illustrates the comprehensive structure and flow of data processing within the system proposed in our work. Initially, the image alignment module, through which input images pass, undergoes training. Subsequently, all feature vectors derived from the training data of the anomaly detection system are enumerated. The pivotal feature vectors are then selectively sampled and stored in the memory bank. This process is an integral aspect of the training procedure for the Abnormality Scoring Module. Following successful completion of the two-stage training, the system becomes primed for the detection of anomalies in chest X-ray images.

For the anomaly detection inference elucidated earlier, the approach put forth in this paper necessitates a two-stage model training procedure. In order to accomplish this, the complete dataset is partitioned into three distinct subcategories in the following manner:
- Image alignment module training dataset
- Anomaly detection system training dataset
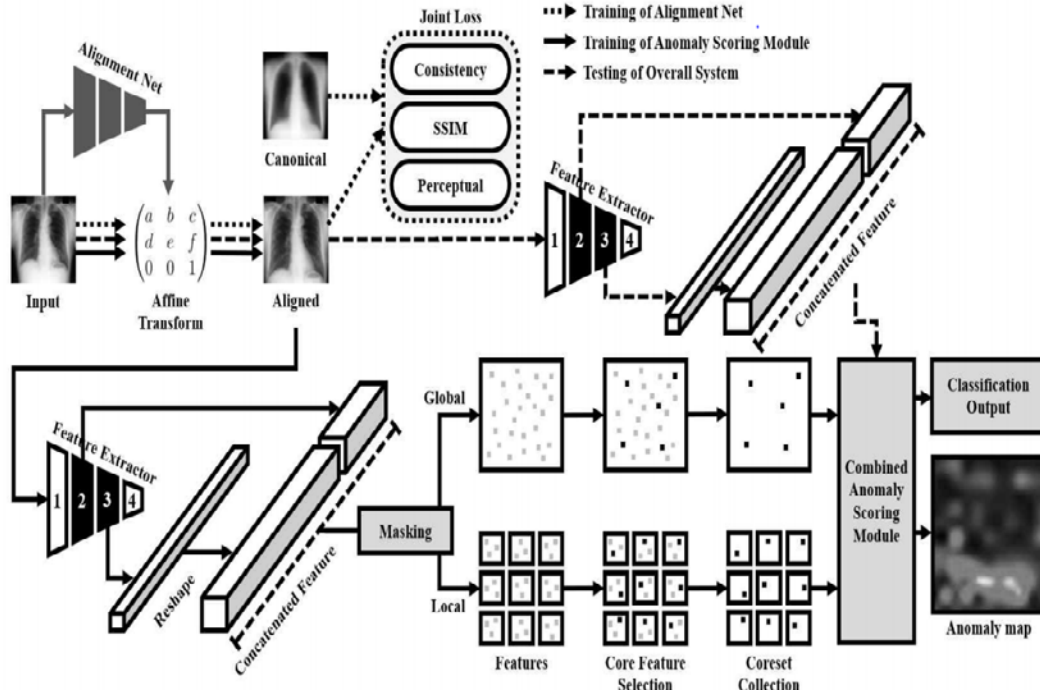- Anomaly detection system experimentation dataset

Fig. 2. Overall schematic diagram for anomaly detection from Chest X-ray images

Within this study, the anomaly detection mechanism leverages a pre-trained deep convolutional neural network model. The convolutional layers of this neural network model, trained on the ImageNet classification dataset [15], enable the extraction of pertinent features for the training task. The extraction of feature maps from the convolutional layers of the pre-trained convolutional neural network is facilitated. The process of extracting feature maps from the employed Wide-ResNet-50 [16] model is shown in **Fig. 3**.
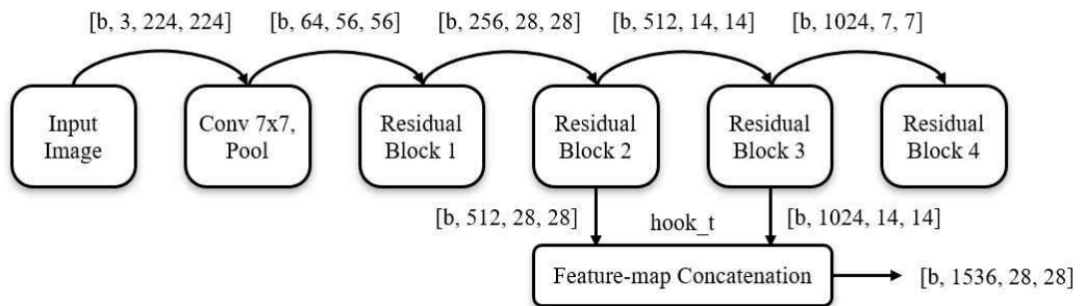

Fig. 3. The process of extracting feature maps from the Wide-ResNet-50 neural network model

Each pixel's feature map is translated into an embedding vector, encapsulating the attributes of the corresponding region in the image. These region-specific feature vectors are termed as patch feature vectors [12]. By contrasting these feature vectors derived from the training dataset with the nearest neighbor vectors acquired via a nearest neighbor search algorithm, it becomes feasible to compute anomalies (referred to as Anomaly Score) [14]. The vector comparison employs a distance function, and the calculation of image-level anomalies

employing the Euclidean distance can be represented as illustrated in Equation (1)

$$d(y) = \frac{1}{K}\sum_{f \in N_k(f_y)}\left\| f - f_y \right\|^2 \tag{1}$$

In Equation (1), the symbol f denotes the feature extractor based on an artificial neural network, N stands for the nearest neighbor search function, and K represents the count of neighbors. Employing this approach to patch feature vectors has shown to enhance the performance of anomaly detection [13]. Nevertheless, a direct application of this technique results in a notable increase in inference time due to the considerable computational overhead associated with neighbor search computations.

To tackle this challenge, the utilization of Core-set Sub-sampling [13, 17] emerges as a solution. This method effectively reduces the computational burden while still harnessing pertinent feature vectors, thereby yielding a marked enhancement in the overall generalization performance of the anomaly detection system. Through the application of Core-set Sub-sampling, the computational workload is streamlined, focusing solely on pertinent and influential feature vectors. This, in turn, contributes substantially to bolstering the system's generalization performance in anomaly detection.

3.1 X-ray Image Alignment

Generally, the PA (Posterior-Anterior) approach for obtaining chest X-ray images follows a consistent acquisition process. Nevertheless, the resulting X-ray images may exhibit variations in structural alignment. The proposed region-specific PatchCore method operates on the premise that extracted local patch features originate from identical anatomical regions within the sub-core sets. This assumption necessitates proper alignment of the input image's perspective, implying a need for standardization or homogenization. Illustrated in **Fig. 4**, the sample images deviate from the canonical image. Thus, there arises a requirement to align the sample images with a reference standard.
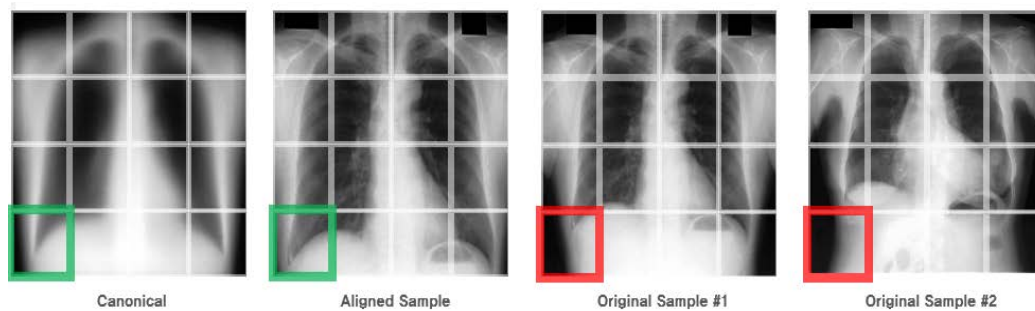

Fig. 4. Comparison of X-ray images: Canonical, Aligned, Original sample x-ray images (from left to right)

The image alignment methods can be classified into following three categories including our proposed scheme as illustrated in **Fig. 5**.

- Searching transform matrix by using iterative image transformation.
- Alignment by detecting landmarks.
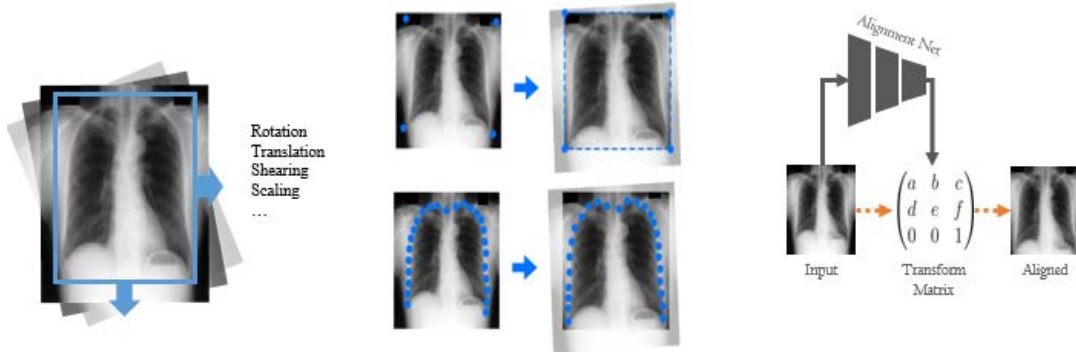- Alignment matrix prediction based on the Alignment Net.

Fig. 5. Image alignment methods
(from left to right, use iterative image transformation, landmarks and Alignment net)

In this investigation, the seamless operation of the Region-Specific PatchCore system is enhanced through the resolution of the inherent challenge of inter-image perspective disparities. This is achieved by employing an automatic image alignment technique as a preprocessing step. The image alignment method utilized in this study draws from a contrastive learning-based approach for thoracic lesion diagnostics [4]. In this scheme, a reference image is synthesized to harmonize the perspective of all thoracic images. This reference image is fashioned by aggregating the luminance values from numerous chest X-ray images that underwent meticulous manual alignment. This resultant image is denoted as the "Canonical Chest Image".

In the context of this research, a randomized subset of 500 chest X-ray images from the dataset underwent direct alignment. This was achieved through a series of transformations including rotation, perspective adjustments, translation, scaling, and shearing, employing bi-cubic interpolation. This sequence yielded a standardized chest X-ray image, referred to as the "Canonical Chest Image," produced by averaging the aligned images. The generated Canonical X-ray image is visualized in **Fig. 6**.
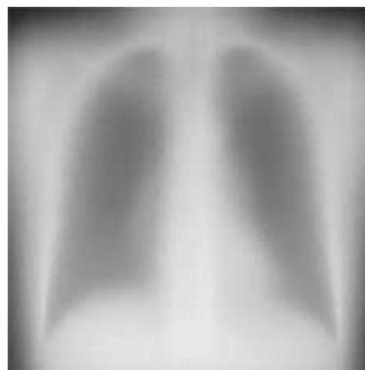

Fig. 6. Generated Canonical X-ray image

After obtaining the Canonical Chest Image, an artificial neural network is trained to predict the transformation matrix that will deform the input images to have a similar perspective as the generated Canonical Chest Image. In the equations described in this section, $\Phi_A(x)$ represents the transformation matrix parameters predicted by the neural network for alignment

purposes. In this study, ResNet-18, ResNet-34, and ResNet-50 [18] were experimented as the artificial neural networks corresponding to Φ_A. The neural network operates as a multi-output regressor that takes the grayscale image X as input data and predicts six affine transformation parameters, excluding homogeneous coordinates, as shown in Equation (2)

$$\Phi_A(x) = \begin{pmatrix} a_i & b_i & c_i \\ d_i & e_i & f_i \end{pmatrix} \tag{2}$$

Consequently, the parameters of an affine transform matrix are evaluated for aligned images using the canonical and sample images. In the proposed method, the neural network Φ_A must be trained to have input images that resemble the structure of the Canonical Chest Image. To achieve this goal, Φ_A utilizes the following three distance functions as loss functions to measure the similarity between the two images: Constancy loss function [19], Reversed SSIM loss function, and VGG perceptual loss function [20] as shown in **Fig. 7**.
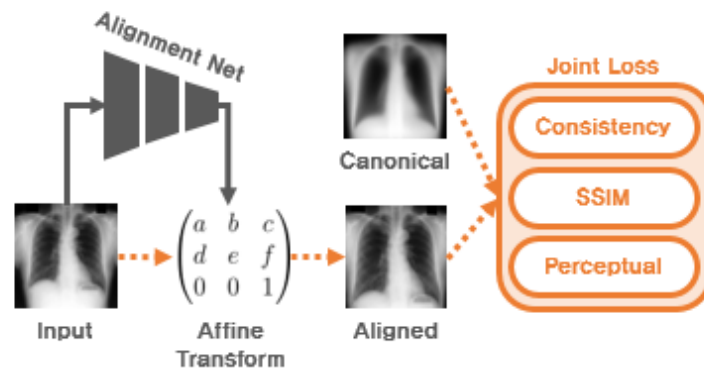

Fig. 7. The process of X-ray image alignment based on the proposed approach

In training for X-ray image alignment, Affine transform can cause gradient vanishing (an image becomes black) and eventually stop the training without decreasing loss value. The sorting network $f(x)$ functions as a regression predictor, and therefore, it performs all pooling operations using average pooling. In addition, it applies the sigmoid activation and is trained with deactivated biases. Unlike the conventional network training methods, when it produces excessively incorrect predictions, it faces a problem where the optimizer loses direction. This problem arises from all affine-transformed images being output as black blank images. It leads to the issue of weights not being updated even as the training progresses. Because, randomly predicted initial parameters of affine transform can cause gradient vanishing.

The goal of the initialization of the predicted parameter is to approximate to the invariant Affine matrix. Therefore, we apply a method to initialize the initial prediction as the base matrix for affine transformations, ensuring that the weights are updated properly. First, we aggregate the initial predictions for the entire dataset immediately after initializing the network with the optimizer deactivated, as shown in Equation (3)

$$\underset{epoch=0}{\mathrm{Y}} = \left\{ \underset{epoch=0}{\mathrm{f}(x)} \middle| \forall x \epsilon X \right\} \tag{3}$$

The aggregated initial predictions, denoted as Y_0, are used to explore the maximum and minimum values. Subsequently, we apply min-max normalization to all network output values. However, excessive standard deviation could lead to the same problem of the optimizer not functioning correctly. To cope with this problem, we utilize the weight α to reduce the prediction variance. Then, as shown in Equation (4), we element-wise sum the base transformation matrices and combine them with the homogeneous matrix to utilize them as affine transformation matrices.

$$
\begin{matrix} n > 1, n\epsilon\Pi \\ \Phi_A(x_i) \\ epcho = n \end{matrix} = \left(\frac{\Phi_A(x_i) \ominus min(Y_0)}{n \atop (max(Y_0) \ominus min(Y_0)) \otimes \alpha} \oplus \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right) \frown (0 \quad 0 \quad 1) \tag{4}
$$

3.2 Region Specific Anomaly Detection

In this study, we propose a method to enhance the efficiency and performance of the existing process through a cascading effect achieved by image alignment, as shown in **Fig. 8**. First, as previously mentioned, the PatchCore based on simple grid partitioning requires images to be aligned with a uniform perspective. Assuming that image alignment is possible, aligned image data allows for the hard masking of unnecessary regions in the extracted feature maps. Feature map hard masking resolves the excessive memory occupation issue commonly encountered in the conventional PatchCore and region-specific PatchCore, significantly reducing the false positive rate.
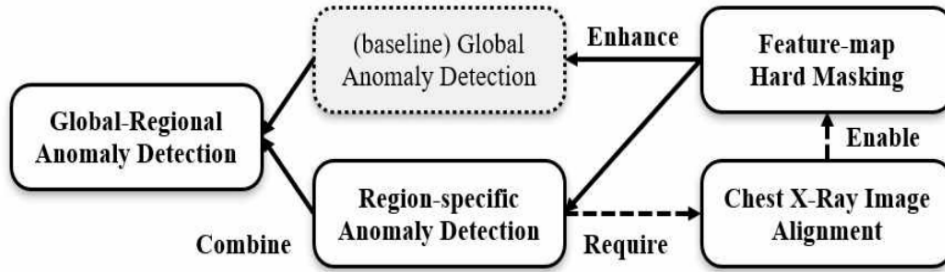

Fig. 8. Chain of region-specific anomaly detection application

The currently available global PatchCore identifies the core vector based on the distance between vectors and does not consider local information in the training phase. The proposed Region specific PatchCore identify the normality based on the local similarity in the training phase like [21]. In this study, we adopt PatchCore as a baseline for anomaly detection. We employ a Wide ResNet 50 pre-trained on the ImageNet classification task as the feature extractor. Only unbiased intermediate-level feature maps among the extracted features are used. When using all the extracted feature data from the entire dataset, it becomes very large in size and susceptible to outliers. Therefore, we subsample the core features to create a representative core set of normal data. During the anomaly score measurement step, we calculate distances to a random number of adjacent normal vectors and take the average value. **Fig. 9** illustrates an exemplary process of region-specific anomaly detection.
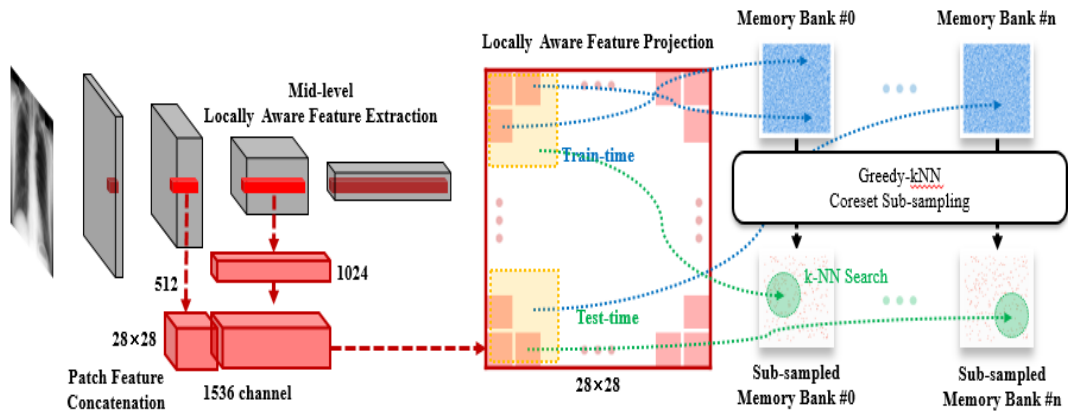
Fig. 9. The process of region-specific anomaly detection

To obtain the region-specific anomaly detection, the input images are divided into n x n feature maps and the Core-set subsampling is performed for each sub image. In this work, we consider n=4. **Fig. 10** illustrates the Core-set subsampling process for the global and region-specific anomaly detection. The combined anomaly score calculation module performs image fusion on the results of the global PatchCore and the region-specific PatchCore heat maps at arbitrary ratios.
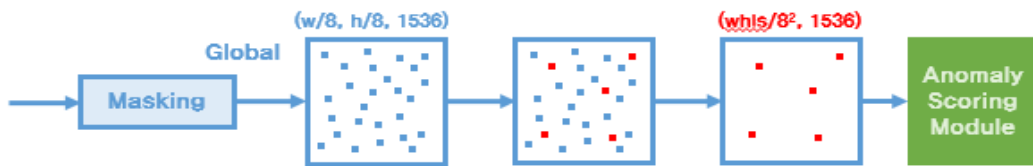
(a) Core-set subsampling for the global PatchCore (baseline approach)
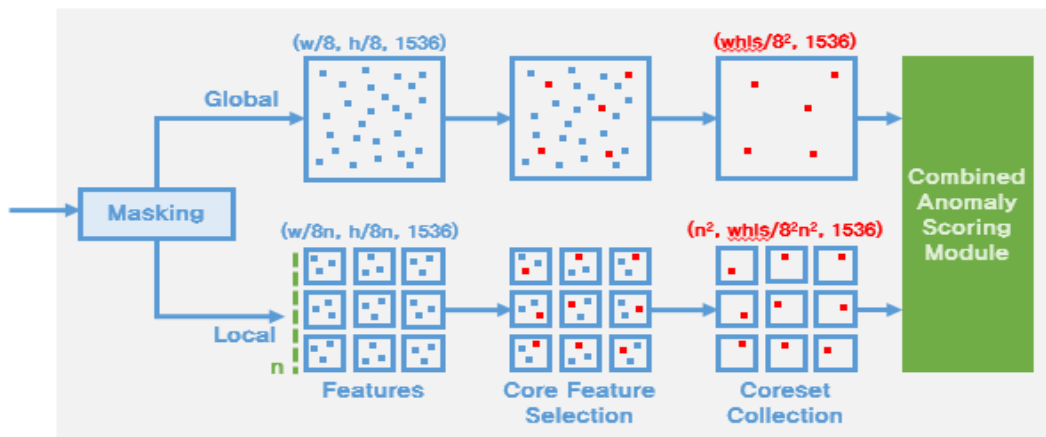
(b) Core-set subsampling for the global and region-specific PathchCore

Fig. 10. The comparison of Core-set subsampling between (a) the baseline approach and (b) the proposed approach (w: width of input image, h: height of input image, n: side length of local features, l: length of total train dataset, s: Core-set subsampling rate)

3.3 Patch Feature Hard Masking

The parallel processing of the global and region-specific anomaly detection needs huge memory and computational burden compared to global only PatchCore method. In this work, in order to improve the memory efficiency in training and computational efficiency in reasoning phase, we utilize Hard Masking by applying the anomaly detected area to extracted feature as shown in **Fig. 11**.

The proposed pipeline in this paper performs image alignment as a preprocessing step just before conducting the anomaly detection process. The image alignment process generates standardized images in which all input chest X-ray images have a consistent orientation. The extracted patch features, inputted with 224×224 resolution images, consist of patch features with sizes of 28×28×512 and 14×14×1024 for each image. The pixels of each patch feature map are represented in 32-bit floating-point format and merging two different-sized patch features results in a size of 28×28×1536. As a result, approximately 4.6MB of feature information is extracted for each image data. In this study, the number of training data utilized for experiments is 1777 images, and it requires approximately 8GB of memory allocation space for the extracted feature data alone. Additionally, to perform vector-to-vector Euclidean distance measurements during the training process, additional memory allocation space is needed, which can increase the memory occupancy rate up to a maximum of 40.3GB. The memory allocation space may further increase proportionally with the input resolution and dataset size. In the existing studies, the proposed method allowed adjusting the trade-off between performance and cost based on the requirements of the usage environment by setting the input resolution and Core-sets subsampling ratio. When increasing the resolution of input images or the Core-sets subsampling ratio, the computational time during training increases proportionally. This implies that even when setting the subsampling ratio to a level that ensures minimal performance based on the trade-off, it still exhibits significantly higher memory occupancy compared to other techniques.

In this study, we propose a scheme to improve both memory efficiency during the training process and computational efficiency during the inference stage by applying hard masking to the features extracted for anomaly detection in the valid regions. The shape of the mask is generated by comparing the average image of the standard images and aligned image data directly. To utilize the necessary chest area from the x-ray image while detecting anomaly, the binary mask is used to remove the areas under a certain intensity level, 10 in our case, based on the average value of 500 x-ray images. The hard making can improve the nearest neighbor searching speed in evaluating a large numbers of feature vectors. We can reduce about 21% (620 pixel/ 784 pixel) of feature vectors which are unnecessary for the process in the experiments.
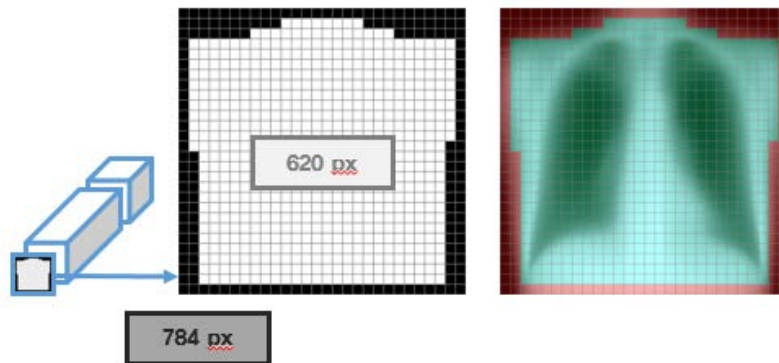

Fig. 11. An example of feature map hard masking: binary mask(left) and feature map mask(right)

Applying hard masking to the extracted features can improve the speed of nearest-neighbor search when dealing with many feature vectors. This reduction in the total number of vectors decreases the computational workload involved in pair-wise distance operations. Furthermore, as only the valid regions corresponding to the actual anatomy are extracted, areas that do not need consideration for normality are excluded, leading to improved anomaly detection performance.

The described feature map hard masking assumes that the input images are aligned, and the extracted feature maps have locality. To perform regional specificity anomaly detection, the input image alignment process was conducted as a preliminary step in this study. Therefore, by applying hard masking, it was possible to enhance the overall performance without the need for a trade-off.

4. Experimental Results

In our experimental setup, we employ an X-Ray Radiograph dataset comprising grayscale single-channel images. Unlike conventional digital color images characterized by 24-bit RGB spread across three channels, X-Ray Radiograph images depict radiation attenuation projected onto the image and primarily adhere to the DICOM standard, spanning a 12-bit range. The dataset utilized in this study encompasses the PadChest [22] dataset, encompassing around 160,000 images in a 16-bit format. To ensure uniformity in the imaging procedure for training and evaluating the anomaly detection system, we exclusively extract and employ images acquired in the PA (Posterior-Anterior) mode.

For the experiments, we utilize 1,777 normal training data and 2,364 testing X-ray images (comprising 1,182 normal and 1,182 abnormal cases) from the PadChest dataset. The assessment is conducted using AUROC (Area Under Curve-Receiver Operating Characteristic). Our development environment encompasses a CPU (Intel i9-10980XE 18c36t), GPU (Nvidia Geforce RTX 3090), RAM (128GB), OS (Ubuntu 20.04), Programming Language (Python 3.8), Deep Learning Framework (PyTorch 1.8.1), CUDA (CUDA 11.1, cuDNN 8.0), and faiss-gpu. In this work, for the purpose of comparing experimental results, we measure the AUROC (Area Under Curve–Receiver Operating Characteristic) for binary classification outcomes. AUROC needs an equal number of normal and abnormal data to be properly evaluated. Typically, it is expressed within the range of 0.5 to 1.0, where values above 0.5 indicate better-than-random performance.

Figs. 12-15 showcase the comparative AUROC results across various conditions, including the utilization of image alignments, hard masking, and a combination of region and global PatchCore.
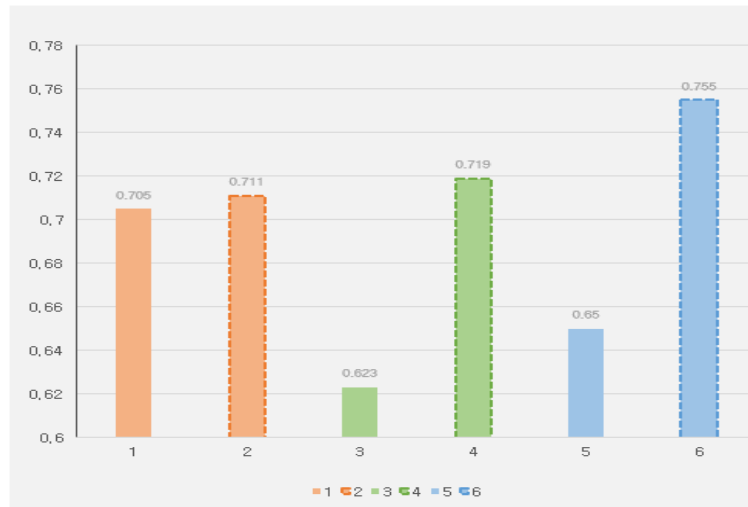
Fig. 12. Comparative results of anomaly detection according to either image aligned or not (1. Not/ PatchCore, 2. Aligned/ PatchCore, 3. Not /Region-specific PatchCore, 4. Aligned/ Region-specific PatchCore, 5. Not/ Combined, 6. Aligned/ Combined)

Fig. 12 illustrates the experimental results regarding the impact of alignment application. While the conventional PatchCore showed minimal performance variation due to alignment, the region-specific PatchCore proposed in this study exhibits a significant disparity based on alignment, highlighting a substantial reliance on alignment conditions. It becomes evident that this region-specific PatchCore on aligned data actually achieves a higher AUROC compared to the traditional PatchCore.
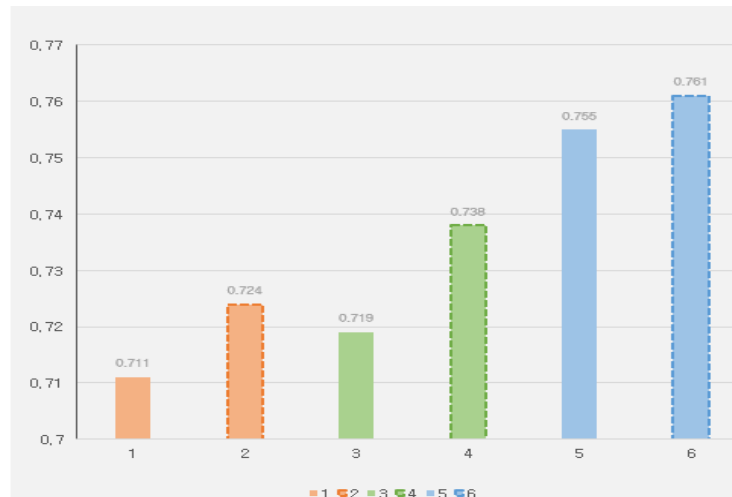

Fig. 13. Comparative result of anomaly detection according to either patch feature hard masking or not (1. Not/ PatchCore, 2. Masking/ PatchCore, 3. Not /Region-specific PatchCore, 4. Masking/ Region-specific PatchCore, 5. Not/ Combined, 6. Masking/ Combined)

Fig. 13 presents a performance comparison based on the application of hard masking to feature maps. Across all experiments, it is evident that hard masking consistently brings about a significant improvement in performance.
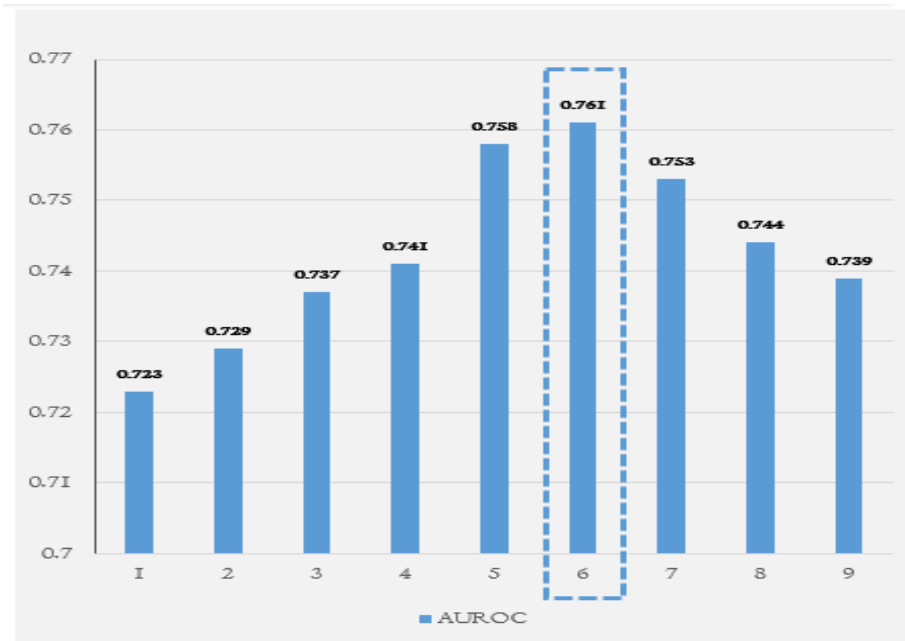
Fig. 14. Performance comparison based on combined heatmap weights of Region-Specific and Global PatchCore

Fig. 14 presents the experimental results conducted to determine the optimal synthesis of prediction heatmap from region-specific and traditional global PatchCore, based on varying weights. We experimented with 9 cases of ratios ranging from 0.1:0.9 to 0.9:0.1 using heatmaps weighted by local and global PatchCore, respectively. The experimental findings indicate that the highest accuracy is achieved (AUROC=0.761) when assigning a weight ratio of approximately 6:4 in favor of region-specific over global contributions. While the exact 6:4 ratio may not be the ultimate optimal, it was observed that a value close to this ratio yields favorable results.

Fig. 15 illustrates the performance comparison with existing methods. The proposed approach under the experimental conditions achieved an AUROC of 0.761. It demonstrated a relatively improved outcome compared to the conventional methods. Furthermore, it was observed that by increasing the input resolution from 224 to 448 and setting the sub-sampling load threshold to 0.01, the performance enhanced to 0.774.
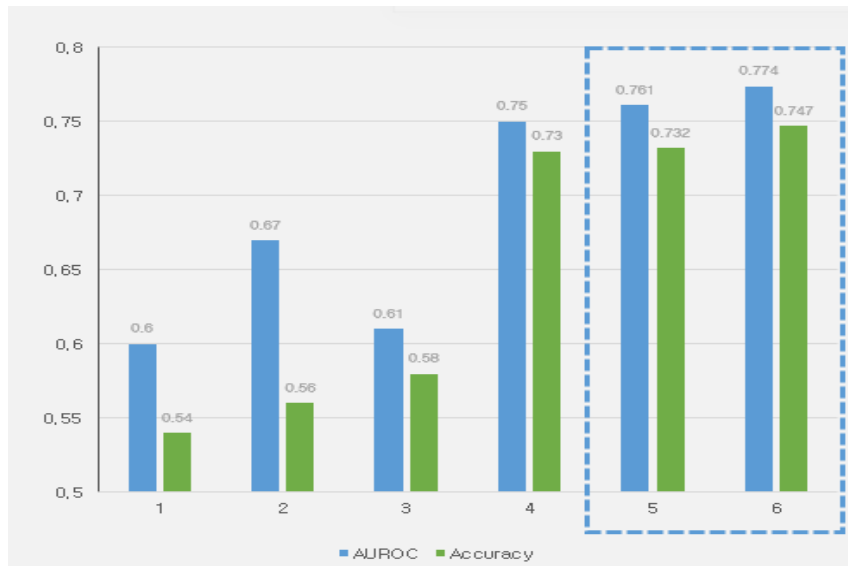
Fig. 15. Comparison of anomaly detection between the existing methods and the proposed method
(1. F-AnoGAN [23], 2. Patch-SVDD [21], 3. Pix2pixHD [14], 4. Patch-SVDD MR Ensemble [21],
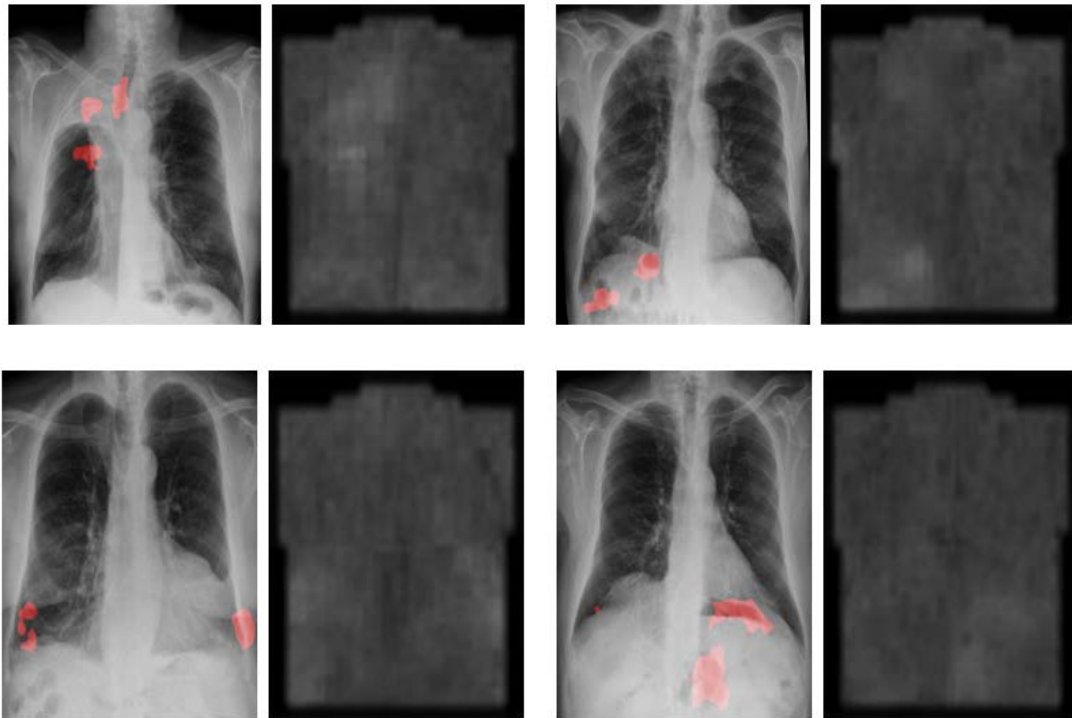5. Proposed Combined (45GB), 6. Proposed Combined (128GB)

Fig. 16. Segmentation results of detected anomaly from X-ray images and their corresponding
anomaly heatmap.

5. Conclusion

Existing artificial intelligence applications conventionally require large-scale training and evaluation datasets for supervised learning. These datasets must include medical imaging data and corresponding labels. Label data varies depending on the task goals of medical artificial intelligence, such as diagnosing the presence of lesions, detecting lesion locations, or predicting the progression of lesions. Generating such label data typically requires analysis from medical experts. This aspect contributes to the difficulty in constructing medical datasets. Furthermore, while it is easy to obtain data for frequently occurring lesions, it can be very challenging to acquire data for rare diseases. This can make it difficult to utilize supervised learning-based classifiers or detectors.

In this research, we aimed to diagnose the presence of lesions using semi-supervised detection method by proposing a method for diagnosing the presence of lesions in chest X-ray images captured in the PA (Posterior-Anterior) projection. Anomaly detection can be utilized effectively to address the challenges encountered in the process of creating datasets for medical artificial intelligence. The task of diagnosing the presence of lesions, which is the focus of this study, shares similar requirements with other tasks where anomaly detection is applied. Since the primary task is to diagnose the presence of lesions, it is more convenient to define the category of normal bodily images rather than specifying numerous characteristics of lesions. This facilitates identifying data that falls outside this category as lesion data.

The approach is based on applying the PatchCore method, which has been effectively utilized for anomaly detection in industrial data, to medical images. The diagnosis of lesion presence in chest X-ray images is established using training data composed solely of normal data. As a result, it outputs binary classification results for normal or abnormal cases. Additionally, during this process, heatmaps recording anomaly scores for each patch are generated, enabling image segmentation.

In order to improve the performance of current anomaly detection systems, we introduce a technique that can identify anomalies specific to certain regions. The method we propose establishes the baseline by segmenting feature vectors extracted from individual image regions. To achieve this, standardization of image composition is necessary, and we recommend a semi-supervised learning approach to align chest X-ray images. Additionally, we put forth a rigorous masking method for aligned images. This method enhances memory utilization, computational speed, and the precision of both established and newly introduced systems.

Through our experiments, the method demonstrates enhanced application efficiency and performance by employing feature map hard masking. Notably, our proposed approach achieves a maximum AUROC of 0.774, showcasing a 6.9% performance improvement over currently available methods that utilize the same dataset.

In this study, images obtained through the PA (Posterior-Anterior) imaging technique were only utilized. However, as for the further work, it will be necessary to expand the application to X-ray images such as such as AP (Anterior-Posterior), bilateral (Left/Right Lateral), Decubitus, Oblique, etc.

References

[1] Tang, Y.-X. et al., "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *npj Digital Medicine*, vol.3, no.1, pp.1-8, 2020. Article(CrossRefLink)

[2] Visuña, L. et al., "Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning," *BMC Medical Imaging*, vol.22, no.1, pp.1-16, 2022. Article(CrossRefLink)

[3] Ligueran, R. J. S.D. et al., "Applied Computer Vision on 2-Dimensional Lung X-Ray Images for Assisted Medical Diagnosis of Pneumonia," *International Journal of Computing Sciences Research (IJCSR)*, vol.7, pp.1239-1254, 2022. Article(CrossRefLink)

[4] Liu, J. et al., "Align, Attend and Locate: Chest X-ray Diagnosis via Contrast Induced Attention Network With Limited Supervision," in *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.10631-10640, 2019. Article(CrossRefLink)

[5] Park, S. et al., "Self-evolving vision transformer for chest X-ray diagnosis through knowledge distillation," *Nature Communications*, vol.13, no.1, pp.1-11, 2022. Article(CrossRefLink)

[6] Hyun-bin Kim and Jun-Chul Chun, "Leision Detection in Chest X-ray Images based on Coreset of Patch Feature," *Journal of Internet Computing and Services*, vol.23, no.3, pp.35-45, 2022. Article(CrossRefLink)

[7] Bergmann, P. et al., "MVTec AD — A Comprehensive Real-world Dataset for Unsupervised Anomaly Detection," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.9584-9592, 2019. Article(CrossRefLink)

[8] Salah, R., Vincent, P., Muller, X., Glorot, X., Bengio, Y., "Contractive Auto-Encoders: Explicit Invariance During Feature Extraction," in *Proc. of ICML'11: Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp.833-840, 2011. Article(CrossRefLink)

[9] Schlegl T. et al., "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," in *Proc. of International Conference on Information Processing in Medical Imaging (IPMI), Springer, Cham*, pp.146-157, 2017. Article(CrossRefLink)

[10] Ruff, L. et al., "Deep One-class Classification," in *Proc. of 35th International Conference on Machine Learning, PMLR*, vol.80, pp.4393-4402, 2018. Article(CrossRefLink)

[11] Bergman, L., Niv, C., and Yedid, H., "Deep Nearest Neighbor Anomaly Detection," *arXiv:2002.10445*, 2020. Article(CrossRefLink)

[12] Defard, T. et al., "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," in *Proc. of Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021, Springer, Cham*, pp.475-489, 2021. Article(CrossRefLink)

[13] Roth, K. et al., "Towards Total Recall in Industrial Anomaly Detection," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.14298-14308, 2022. Article(CrossRefLink)

[14] Cohen, N., Yedid H., "Sub-Image Anomaly Detection with Deep Pyramid Correspondences," *arXiv:2005.02357*, 2020. Article(CrossRefLink)

[15] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., "ImageNet: A large-scale hierarchical image database," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.248-255, 2009. Article(CrossRefLink)

[16] Zagoruyko, S., Komodakis, N., "Wide Residual Networks," *arXiv:1605.07146*, 2016. Article(CrossRefLink)

[17] Sener, O., Savarese, S., "Active Learning for Convolutional Neural Networks: A Core-Set Approach," in *Proc. of 6th International Conference on Learning Representations (ICLR 2018)*, 2018. Article(CrossRefLink)

[18] He, K. et al., "Deep Residual Learning for Image Recognition," in *Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778, 2016. Article(CrossRefLink)

[19] Zhu, J.-Y. et al., "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proc. of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp.2242-2251, 2017. Article(CrossRefLink)

[20] Ledig, C. et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.105-114, 2017. Article(CrossRefLink)

[21] Yi, J., Yoon, S., "Patch SVDD: Patch-Level SVDD for Anomaly Detection and Segmentation," in *Proc. of 15th Asian Conference on Computer Vision – ACCV 2020*, 2021. Article(CrossRefLink)

[22] Bustos, A. et al., "PadChest: A large chest x-ray image dataset with multi-label annotated reports," *Medical Image Analysis*, vol.66, 2020. Article(CrossRefLink)

[23] Schlegl, T. et al., "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Medical Image Analysis*, vol.54, pp.30-44, 2019. Article(CrossRefLink)

Hyunbin Kim is a senior engineer in ADDD (https://addd.co.kr/) at Seoul, South Korea. He has received a B.S. degree, majoring in Computer Science at Kyonggi University. He was also conferred M.S. degrees in the field of Computer Vision & Deep Learning at Kyonggi University. His current research interests include Computer Vision, Deep Learning, and Human Computer Interaction.

Junchul Chun is currently a professor in the department of computer science and a principal investigator in the Graphics and Image Processing Research Laboratory at Kyonggi University, Suwon South Korea. He received a B.S. degree from Chung-Ang University majoring in computer science. He also received M.S. and Ph.D. degrees from the department of computer science and engineering at the University of Connecticut, CT., U.S.A, in 1992 and 1995, respectively. He was a visiting research scholar in PRIP (Pattern Recognition and Image Processing) at Michigan State University and University of Colorado at Boulder, 2001 and 2009 respectively. His major research fields are Augmented Reality, Computer Vision, and Human Computer Interaction.