International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 4
ISSN: 2088-5334

PatchCore-based Anomaly Detection using Major Object Segmentation

Gyu-Il Kim [a], Kyungyong Chung [b,*]

[a] Department of Computer Science, Kyonggi University, Suwon, Republic of Korea
[b] Division of AI Computer Science and Engineering, Kyonggi University, Suwon, Republic of Korea
Corresponding author: *dragonhci@gmail.com

Abstract—**Cameras utilized for product defect detection in the production line of the manufacturing process create noise due to environmental changes such as camera angle and direction of light. This causes a lack of manufacturing process data and reduces the efficiency of anomaly detection. Therefore, it is necessary to produce a method that detects defects occurring in the production line and guarantees product quality and safety using anomaly detection technology combined with artificial intelligence. Therefore, this thesis proposes PatchCore-based anomaly detection using major object segmentation. The proposed method pre-processes product packaging data by using Green Channel thresholding, Major Connected Component Selection, Extracting Outer Contour, and FloodFill with Centroid. As for the pre-processed data main objects are masked, and the image data is segmented. Through PatchCore model, normality and anomaly detection results are binarily classified. In the performance evaluation, the accuracy is compared between the pre-existing anomaly detection method and the proposed method through the pre-/post-preprocessing data, and high performance is proven. The conventional method showed an accuracy of 0.7684, while our approach achieved an accuracy of 0.9784. Additionally, among the CNN models, VGG19 demonstrated an accuracy of 0.5833, and EfficientNet80 showed an accuracy of 0.7, both of which were lower than our method's accuracy. Therefore, even a small data set shows strong performance through the proposed method. The proposed method is expected to be utilized as an effective defect detection model in diverse fields.**

Keywords— **Anomaly detection; deep learning; machine learning; artificial intelligence; image segmentation.**

Manuscript received 10 Dec. 2022; revised 10 Mar. 2023; accepted 17 May 2023. Date of publication 31 Aug. 2023.

I. INTRODUCTION

In the manufacturing process, the field of defect detection combined with artificial intelligence is making rapid progress. In addition, attempts are being made to utilize automated image analysis through image detection equipment such as cameras [1], [2]. In particular, the product packaging process is an important stage for guaranteeing product safety and quality. Defects occurring in the production line of the product packaging process are threats to product safety and quality. Such defects must be detected and removed before final products are packaged and delivered to customers. To do so, anomaly detection technology combined with artificial intelligence plays an important role in the product packaging process. The pre-existing defect detection technology is mainly dependent on human experience and senses. Such a method is uneconomical and may cause human errors during classification. Recently, automated defect detection technologies have been studied along with machine learning and deep learning technologies [3]. Defect detection technology applies to diverse industrial fields. In particular, as far as the industrial field is concerned, defect detection is being commercialized in characterized fields such as defect detection in manufacturing, defect prediction in energy generation, and disease diagnosis in healthcare [4], [5].

Recently, studies are being conducted on defect detection based on anomaly detection [6]. Anomaly detection is a technology used to detect objects or events that malfunction or deviate from the normal patterns of a data set [7], [8]. Anomaly detection is used in diverse fields such as data mining, statistics, machine learning, and automated processes, and is useful for detecting data set anomalies and analyzing the cause of such anomalies. Anomaly detection mostly utilizes machine learning [9]. Machine learning utilizes a normal data set to enable model learning and determines whether new data is normal data or abnormal data. Anomaly detection utilizes unsupervised learning techniques [10]. Since unsupervised learning utilizes only normal data to enable model learning, it is suitable for detecting outliers distinguished from normal data. Anomaly detection is being utilized in diverse fields, such as manufacturing, security, and medicine [11].

In manufacturing, anomaly detection detects anomalies or defects by monitoring the data generated while the automated process is properly performed [12]. Through such a process, anomaly detection plays a role in contributing to quickly detecting anomalies and taking action on a real-time basis [13]. Defective products are discovered through anomaly detection, and this decreases the risk of production suspension, product recall, etc. This process enhances process efficiency. However, anomaly detection experiences many difficulties in the industrial field. There is a data imbalance problem in that the volume of abnormal data is significantly smaller than normal data. In addition, diverse variable elements such as camera angle, the direction of lighting, and the lighting color exist in every environment. Therefore, anomaly detection is limited because it cannot be widely applied to diverse industrial fields [14]. Therefore, developing a widely applicable defect detection method capable of using a small volume of data to detect product defects is necessary.

This study uses machine learning and deep learning technologies, and diverse pre-processing techniques are utilized to perform anomaly detection in the product packaging process. Through this, a small volume of data is utilized to improve accuracy. In addition, utilizing and applying the involved pre-processing techniques to other industrial process lines can be considered. Through this, defects occurring in the production line can be detected in real-time to guarantee product quality and safety. To evaluate the method proposed in this study, an evaluation is performed to evaluate the accuracy of the proposed method. The method proposed in this study makes contributions as follows.

- The proposed pre-processing technique and PatchCore-based anomaly detection can be utilized to resolve the data imbalance problem by obtaining high accuracy with a low volume of data, and the proposed technique and PatchCore based anomaly detection can be applied to actual industrial fields.
- An anomaly detection technique that combines diverse pre-processing methods can be utilized and widely applied to industrial fields.
- In manufacturing, it can promote the advancement of anomaly detection technology and enhance product safety and quality.

II. MATERIALS AND METHOD

As for most manufacturing process data, the volume of abnormal data is significantly smaller than that of normal data. In addition, diverse variable elements such as camera angle, light direction, and light color exist in every field [15]. Therefore, improving such data's variable elements enables performance improvements. To do so, it is necessary to produce procedures for identifying the data types and utilizing adequate pre-processing methods.

The process of PatchCore-based anomaly detection using major object segmentation proposed in this study is as shown in Fig. 1. For performance improvement purposes, the constructed method mainly consists of two stages: Major Object Area Segmentation and Create Segmented Area Image. In the first stage, main objects are segmented and masked through the stages of Green Channel thresholding [16], Major Connected Component Selection [17], Extracting Outer Contour [18], and FloodFill with Centroid [19]. In the second stage, the segmentation of image data is applied based on the regions masked in the first stage. Therefore, since a PatchCore-based algorithm receives the segmented images, it operates more effectively and obtains higher accuracy.

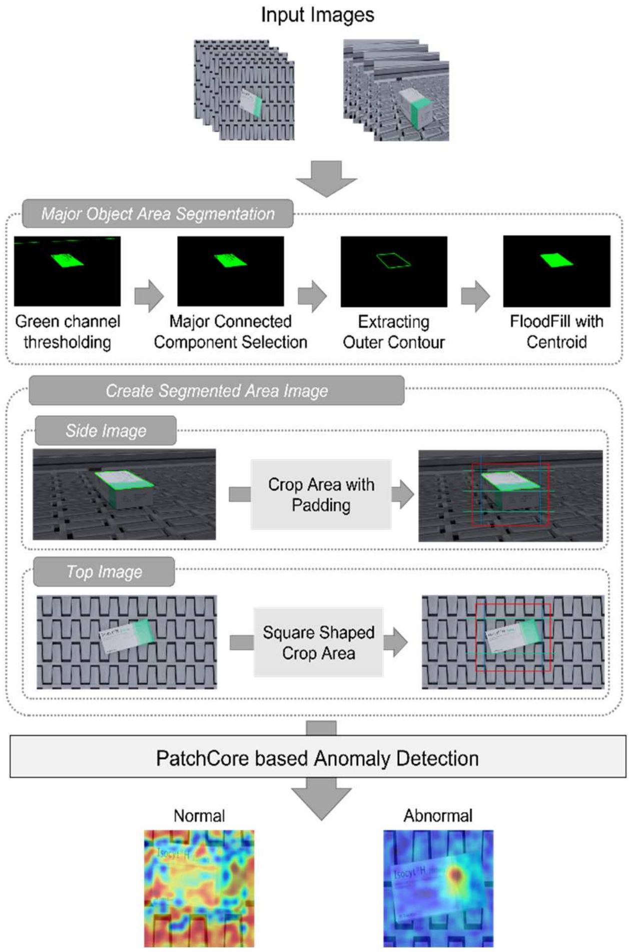
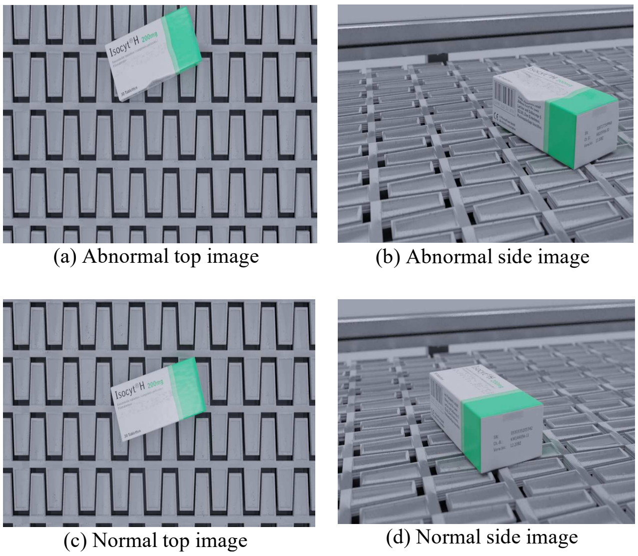


Fig. 1 Process of PatchCore-based Anomaly Detection using Major Object Segmentation

A. Major Object Area Segmentation

The data used in this study is actual industrial field manufacturing process data. It consists of images of drug packaging boxes recorded on a conveyor belt. As for the data, images shot from two directions (product side view and top view images) constitute a pair. The data consists of a total of 400 images consisting of 100 pairs of normal product images and 100 pairs of abnormal product images. Each pair consists of a side view image and a top view image and can be used in one-to-one correspondence. Fig. 2 shows side-view and top-view images of normal data and abnormal data. In Fig. 2, it can be confirmed that the main color can be divided between the conveyor belt and the product. This exists as a difference in the color reflectivity of the conveyor belt and product. Through this, the main color ingredients per region of an image can be used to segment the object [20]. Fig. 3 visually shows the anomaly detection results obtained from the original data using a general model.



(a) Abnormal top image (b) Abnormal side image

(c) Normal top image (d) Normal side image

Fig. 2 Normal and Anomaly Images

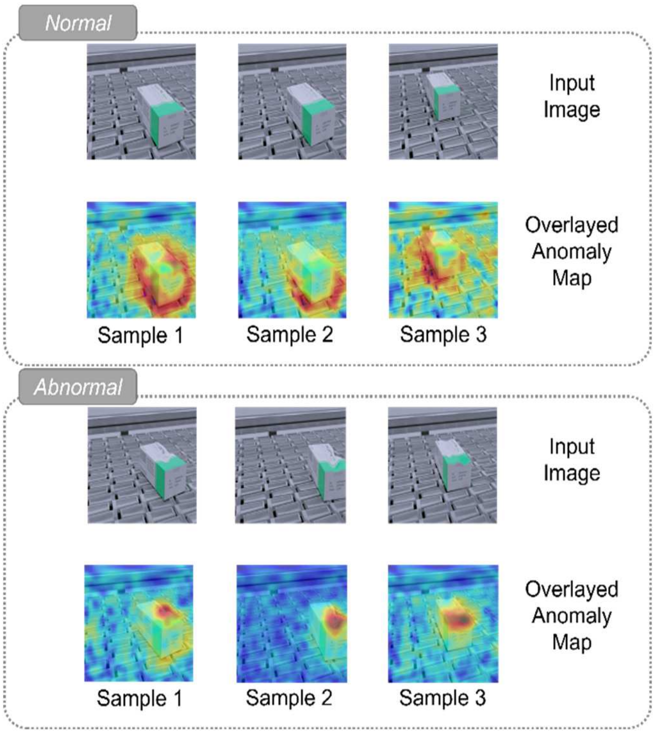


Fig. 3 Anomaly Heat Map of PatchCore with original data

Fig. 3, it can be confirmed that abnormal anomalies existing in abnormal data are well detected. However, it can be confirmed that anomalies existing in normal data are detected as well. What is causing such results are as follows. First, learning the shade patterns generated underneath an object from a small data set is difficult. Second, considering the object's size, the image consists of excessively wide regions. To resolve this problem, separating the background by segmenting the main object is necessary. Fig. 4 shows the main color ingredient results per region of the image.

Fig. 4 Visualization of Major Color Component

In Fig. 4, the image's color difference can be confirmed. Through this, basic segmentation is performed by thresholding the brightness value to adjust to the brightness value of the data. In this study, of the channels that construct the RGB color model, only the G channel's brightness value is thresholded to 180. Then, the conveyor belt existing in the upper section of the image is partially segmented as well. To remove this, the main connected regions are selected. Then, to remove the label regions, such as the product name printed on the inside of the product, it is necessary to perform morphological region operation [21] and outer contour extraction. Morphological region operation may cause the following problems: not recording the entire product and damaging mask precision in a cropped image. Therefore, in this study, the outer contour extraction method is used. Then, a flood fill algorithm fills the inner region color along the outer contour. Fig. 5 shows the pre-processing process in which the upper area of the box serving as the main object is segmented from the original product image.

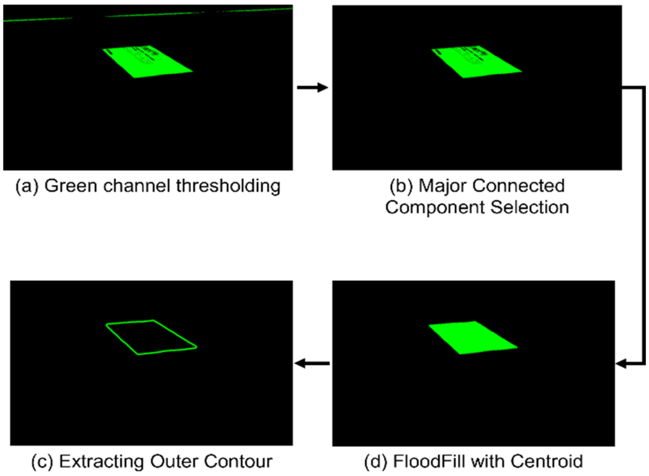
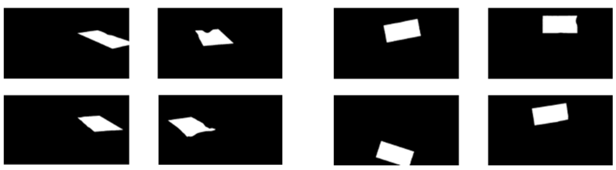


Fig. 5 Result of each pre-processing processes

Fig. 6 shows a masked image of the object divided through the pre-processing process. Fig. 6 (a) shows the side image's segmentation mask, and Fig. 6 (b) shows the top image's segmentation mask. Through Fig. 6, it can be confirmed that the image data enables successful product segmentation.

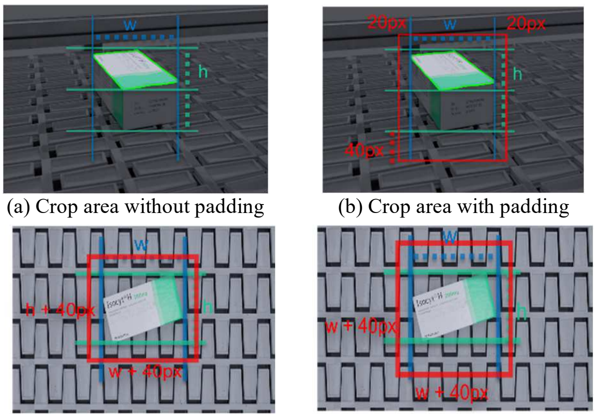


(a) Side Image Segmentation Mask (b) Top Image Segmentation Mask

Fig. 6 Result of segmentation masks

B. Create Segmented Area Image

To improve the performance of PatchCore-based anomaly detection, it is necessary to generate segmented regions. The side image's segmentation mask and the top image's segmentation mask in Fig. 6 are utilized. As for the side image, the region extended downward from the previous mask to a distance equal to the height of the previous mask is segmented. Since the product region can be partially segmented when the segmentation is performed without leaving a margin, the image is segmented by setting a padding of 20px. As for the top image, it must be segmented into square shapes. This is because the image is modified into a square shape, and the ratio is randomly changed in the anomaly detection process. Therefore, the segmentation process segments the top image into square shapes. Fig. 7 shows the segmentation results of the side image and top image.



(a) Crop area without padding (b) Crop area with padding

(c) Rectangular crop area (d) Square shaped crop area

Fig. 7 Result of crop method for side and top product image

When segmentation is applied by utilizing the coordinates of the generated segmentation mask, the final image is as follows. Fig. 8 shows the results of the segmented images.

(a) Cropped side image (b) Cropped top image

Fig. 8 Result of Cropped Images

C. PatchCore based Anomaly Detection

PatchCore-based anomaly detection is one of the technologies used to detect anomalies from visual data such as images [22]. PatchCore is mainly divided into two stages. In the first stage, normal data is utilized to extract and save features through CNN technology [23]. Feature extraction through CNN uses mild-level features instead of high-level features. The subsampling of features extracted from the mid-level is performed [24]. The features reduced through subsampling are saved in the memory bank. In the second stage, the features of new input data are extracted. The distance of the extracted patch features is compared to the distance saved in the memory bank. The anomaly score is measured through the compared distances, and the anomaly status is determined.

PatchCore is set as follows. First, Coreset-Subsampling is performed before saving the extracted features in the memory bank. The Coreset-Subsampling rate is set as 5%. Before the image is input into a model, the image is cropped to have measurements of 224 x 224. The accuracy is measured once the training is completed through 5-fold cross-validation [25]. Fig. 9 shows a process of a PatchCore-based anomaly detection model.

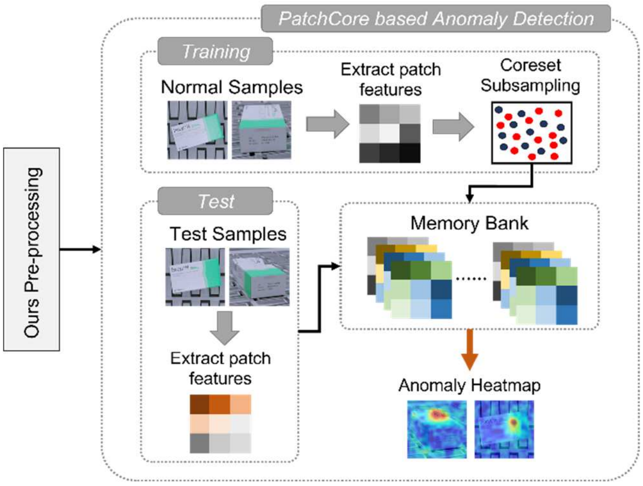


Fig. 9 Process of PatchCore based Anomaly Detection

Fig. 10 and Fig. 11 show the heat map of a normal image and an abnormal image, respectively, extracted through the model.

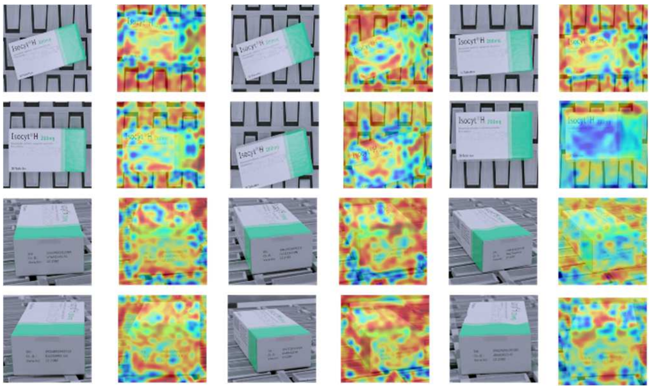


Fig. 10 Heat Map of Normal Sample

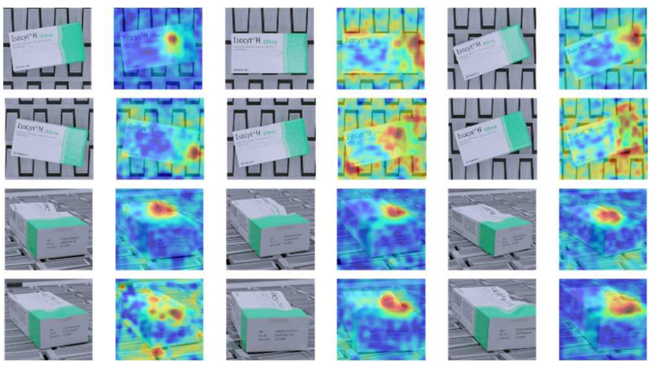


Fig. 11 Heat Map of Abnormal Sample

Unlike the experimental results obtained from the original data, the model could not detect anomalies in the normal image, and the overall heat map appears to be red. This is because a heat map relatively displays outliers and therefore appears to be red overall when no anomalies are detected in a particular area. On the other hand, the model detected anomalies in the abnormal image, and only the involved part appeared to be read. Through this, it can be confirmed that the model detects anomalies only in the abnormal image.

III. Result and Discussion

Data consists of a total of 400 images that form 200 pairs consisting of one side image and one top image each. Anomaly detection is performed for product defect detection, and anomaly detection refers to a machine learning-based binary classification method that is trained using normal data only. Therefore, the model is trained using 100 pairs of normal images.

In the performance evaluation of the binary classification model, the accuracy, precision, and recall resulting from each random threshold value can be measured [26], [27]. As a performance evaluation index that does not have a random threshold value setting, the AUC-ROC (Area Under Curve – Receiver Operating Characteristic) can be utilized [28], [29]. The AUR-ROC is abbreviated and known as AUROC (Area Under the Receiver Operating Characteristic Curve). AUROC refers to the area under the ROC curve, a graph created by plotting the TPR (True Positive Rate) and FPR (False Positive Rate) points that are each specific to a particular threshold. The TPR and FPR are greatly influenced by how the evaluation data is constructed. Therefore, to be able to measure the AUROC of a product defect detection model accurately, the evaluation data set must be constructed to have normal data and abnormal data in a ratio of 1:1. To accurately measure the AUROC of this data set, the evaluation data set was assigned some of the 100 pairs of normal data.

The evaluation data was assigned 20 pairs of normal and abnormal data, and the training data was assigned the remaining 80 pairs of normal data. The hardware consisting of Intel i9-10980XE, 128GB Memory, and NVIDIA GeForce RTX 3090 is used to do so. As for the software, the environment was configured using Python (Ver 3.9.7) and PyTorch (Ver 1.8.1) [30]. The results obtained from the anomaly detection performed using non-preprocessed original data and the results obtained through applying the technique proposed in this thesis are shown in Table 1.

TABLE I
RESULTS OF VARIOUS MODELS

	Category	AUROC
CNN	VGG19	0.5833
	EfficientNet80	0.7
Vanilla PatchCore	Top	0.5923
	Side	0.7871
	Top + Side	0.7684
Ours	Top	0.9098
	Side	0.9750
	Top + Side	**0.9784**

Table 1 shows the experiment results. As for CNN, non-preprocessed original data were learned by VGG19 and EfficientNet80 models. As for Vanilla PatchCore, non-preprocessed original data was learned by a PatchCore based anomaly detection model. The heat map of the original data experiment is shown in Fig. 3.

Table 1 shows the performance of this study. The measured AUROC was 0.9784, indicating that the result is excellent. In addition, it shows the results of a CNN model obtained from non-preprocessed original data. The VGG19 model showed an accuracy of 0.5833, and the EfficientNet80 model showed an accuracy of 0.7. By comparing the CNN and Standard, it can be confirmed that, when the original data is utilized, the PatchCore based anomaly detection model showed higher accuracy than the CNN model. Table 2 shows the accurate results obtained by merging two data types and utilizing one model. The combination is performed by learning the normal top images and normal side images simultaneously.

TABLE II
EXPERIMENTAL RESULTS WHEN MERGING DATA

Category	Accuracy	Precision	Recall
SUM (TOP+SIDE)	0.905	0.846154	0.99

Based on the results shown in Table 2, it can be seen that the recall is increased but that the precision is decreased. Instead of simply merging data, the probability can be drawn through each model, and the final classification results can be drawn through AutoML. Table 3 shows the accuracy results drawn through AutoML. The anomaly score is drawn, and the results are drawn through AutoML.

TABLE III
EXPERIMENTAL RESULTS WHEN MERGING WITH AUTOML

Category	Accuracy	Precision	Recall
SUM (TOP+SIDE AutoML)	0.9255	0.9103	0.9502

Based on the results shown in Table 3, it can be seen that the recall was lower than that achieved through a simple merger, but the precision was increased, and overall accuracy increased.

IV. Conclusion

PatchCore-based anomaly detection is robust against cold-class anomalies and performs effectively even when small data sets are utilized. However, its performance decreases depending on the characteristics of the data. The main reason for this problem is that the surrounding information is also learned as the objects are learned. In this study, to resolve such a problem, diverse pre-processing methods are applied. Through such an application, the accuracy performance was improved from 0.7684 to 0.9784. This shows that the pre-processing methods removed the background and enhanced learning efficiency. Through this, while accepting the advantage that anomaly detection does not define the features of abnormal data, the features of anomaly-free normal data can be defined to detect abnormal data deviating from the defined features. In addition, when performing defect classification using a small data set of fewer than 500 images, a pre-trained deep neural network can be utilized to perform excellent performance in environments involving difficult data volumes. Through this, it is possible to perform effective anomaly detection even when a small data set is used, and this can be utilized in diverse fields. In particular, it is possible to perform effective defect detection even when a small data set consisting of manufacturing processes is used, and it is possible to maintain product quality. Through this, it is possible to perform effective anomaly detection and defect classification regardless of the data set volume or type, and this can be utilized in diverse fields.

In the future, it is scheduled to conduct a study to apply the proposed method to industrial data and other fields such as medical data. In addition, it is scheduled to conduct a study aimed at integrating the pre-existing two stages of region segmentation into one stage by improving the algorithm. In addition, it is scheduled to conduct a study to improve the performance of a PatchCore-based anomaly detection model by changing the deep neural network.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1068828). Additionally, this work was supported by Kyonggi University's Graduate Research Assistantship 2023.

REFERENCES

[1] K. Morimoto, A. Ardelean, M. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Opt. Lett*., vol. 7, no. 4, pp. 346-354, 2020.

[2] Z. Ren, F. Fang, N. Yan, and Y. Wu, "State of the art in defect detection based on machine vision," *IJPEM*., vol. 9, no. 2, pp. 661-691, 2022.

[3] P. Boniol, M. Linardi, F. Roncallo, and T. Palpans, "Automated anomaly detection in large sequences," in *Proc. - Int. Conf. Data Eng*., TX, USA, 2020, pp. 1834-1837.

[4] A. Abid, M. T. Khan, and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artif. Intell. Rev*., vol. 54, pp. 3639-3664, 2021.

[5] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, "Surface defect detection methods for industrial products: A review," *Appl. Sci*., vol. 11, no. 16, pp. 7657, 2021.

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv*., vol. 41, no. 3, pp. 1-58, 2009.

[7] W. Yan and L. Yu, "On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach," in *Proc. Annu. Conf. Prognostics Health Manage. Soc*., vol. 6, 2015. [Online]. Available: https://www.phmsociety.org/ node/1652.

[8] Z. Li, J. Li, Y. Wang, and K. Wang, "A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment," *Int. J. Adv. Manuf. Technol*., vol. 103, pp. 499-510, 2019.

[9] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," *Adv. Eng. Inform*., vol. 45, pp. 101105, 2020.

[10] B. U. Jeon, and K. Chung, "CutPaste-Based Anomaly Detection Model using Multi Scale Feature Extraction in Time Series Streaming Data," *KSII Trans. Internet Inf. Syst*., vol. 16, no. 8, 2022.

[11] J. Seo, J. Park, J. Yoo, and H. Park, "Anomaly Detection System in Mechanical Facility Equipment: Using Long Short-Term Memory Variational Autoencoder", *J Korean Soc Qual Manag*, vol. 49, no. 4, pp. 581-594, 2021.

[12] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*., pp. 9592-9600, 2019.

[13] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *IEEE Access*., vol. 9, pp. 78658-78700, 2021.

[14] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimed. Tools Appl*., vol. 80, pp. 16979-16995, 2021.

[15] J. Dib, K. Sirlantzis, and G. Howells, "A Review on Negative Road Anomaly Detection Methods," *IEEE Access*., vol. 8, pp. 57298-57316, 2020.

[16] V. Thamilarasi, R. Roselin, "Automatic Thresholding for Segmentation in Chest X-Ray Images Based on Green Channel Using Mean and Standard Deviation," *Int. j. eng. innov. technol*., vol. 8, no. 8, pp. 695-699, 2019.

[17] Y. Weng, C. Xia, "A new deep learning-based handwritten character recognition system on mobile computing devices," *Mob. Netw. Appl*, vol. 25, pp. 402-411, 2020.

[18] T. Goel, K. C. Tripathi, and M. L. Sharma, "Single Line License Plate Detection Using OPENCV and tesseract," *Int Res J Eng Technol*, vol. 7, no. 5, pp. 5884-5887, 2020.

[19] A. Roychoudhury, M. Missura, and M. Bennewitz, "Plane segmentation in organized point clouds using flood fill," *Proc. - IEEE Int. Conf. Robot. Autom*., pp. 13532-13538, May. 2021.

[20] M. Chen, T. Artieres, and L. Denoyer, "Unsupervised object segmentation by redrawing", *NeurIPS*., vol. 32, 2019.

[21] C. Chen, S. Yi, J. Mao, F. Wang, B. Zhang, and F. Du, "A Novel Segmentation Recognition Algorithm of Agaricus bisporus Based on Morphology and Iterative Marker-Controlled Watershed Transform," *Agronomy*, vol. 13, no. 2, pp. 347, 2023.

[22] N. Cohen, Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," arXiv preprint arXiv:2005.02357, 2020.

[23] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *1st International Workshop on Industrial Machine Learning*, 2021, pp. 475-489.

[24] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 14318-14328, 2022.

[25] A. Satria, O. S. Sitompul, and H. Mawengkang, "5-Fold Cross Validation on Supporting K-Nearest Neighbour Accuration of Making Consimilar Symptoms Disease Classification," in *Conf. Computer Science and Engineering* 2021, vol. 1, pp. 1-5.

[26] H. Yoo, R. C. Park, and K. Chung, "IoT-based health big-data process technologies: a survey," *KSII Trans. Internet Inf. Syst*., vol. 15, no. 3, pp. 974-992, 2021.

[27] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Appl. Soft Comput*., vol. 86, pp. 105836, 2020.

[28] H. Yoo, K. Chung, "Classification of Multi-Frame Human Motion Using CNN-based Skeleton Extraction," *Intell. Autom. Soft Comput*., vol. 34, no. 1, 2022.

[29] N. T. Mohammad, S. A. Ismail, M. N. kama, O. M. Yusop, and A. Azmi, "Customer churn prediction in telecommunication industry using machine learning classifiers," in *Proc. 3rd Int. Conf. Vis., Image Signal Process*., Aug. 2019, pp. 1–7.

[30] S. Imambi, K. B. Prakash, and G. R. Kanagachidambaresan, "PyTorch," in *Programming with TensorFlow: Solution for Edge Computing Applications*, Cham, Switzerland: Springer, 2021, pp. 87-104.