Enhancing Anomaly Detection Accuracy with Conditional CutPaste

1st Haruki Honda
Department of Robotics Science and Technology
College of Engineering
Chubu University
Aichi, Japan
tr25015-9376@sti.chubu.ac.jp

2nd Yuji Yamauchi
Department of Robotics Science and Technology
College of Engineering
Chubu University
Aichi, Japan
yuu@fsc.chubu.ac.jp

***Abstract*—Recently, anomaly detection using artificial intelligence (AI) has advanced rapidly. A key challenge in anomaly detection is the scarcity of anomalous data for training. To address this issue, a method called CutPaste has been proposed, which involves cutting a region from an image and pasting it elsewhere within the same image. However, since CutPaste performs this operation at random positions, it sometimes generates unnatural anomalies that are unlikely to occur in the real world.**

In this study, we propose an anomaly detection method using conditional CutPaste, which generates more realistic pseudo-anomalous images. Our method utilizes a pair of similar normal images: a region is cut from one image and pasted onto the corresponding region of the other. By training the model with only pseudo-anomalous images that resemble real-world anomalies, we aim to improve classification accuracy. Evaluation experiments conducted on the MVTec AD dataset demonstrated that the proposed method improved accuracy by an average of 4.7% compared to conventional approaches.

***Index Terms*—anomaly detection, cutpaste, deep learning**

I. Introduction

Anomaly detection plays a critical role in various fields, including manufacturing, healthcare, and infrastructure monitoring. However, building effective anomaly detection models presents a significant challenge—namely, the difficulty of collecting anomalous data for training. Due to the nature of the problem, anomalous instances are extremely outnumbered by normal ones.

To address this issue, a method called CutPaste [1] has been proposed. CutPaste generates pseudo-anomalous images by randomly cutting a patch from a normal image and pasting it onto a different location within the same image. It has been reported that training with such artificially generated anomalies can lead to high anomaly detection performance. However, because patches are pasted at random locations, CutPaste often produces unrealistic anomalies that are unlikely to occur in real-world scenarios. These unnatural anomalies can hinder the performance of the detection model.

To overcome this limitation, we propose conditional CutPaste, a method for generating pseudo-anomalies that are more realistic and subtle, similar to those that may actually occur in practice. Our method uses two normal images with similar poses: a patch is cut from one image and pasted onto the corresponding region of the other. This generates pseudo-anomalous images that more closely resemble real anomalies. We adopt ResNet-18 [2] as the base model and formulate the task as a binary classification problem between normal and anomalous classes. By training the model with more natural pseudo-anomalies, we aim to improve classification accuracy.

II. Proposed Method

An overview of the proposed method is illustrated in Figure 1. As shown in Fig. 1(a), two normal images are randomly selected from the training dataset, which consists solely of normal images, and the similarity between them is computed. To evaluate similarity, we use Structural SIMilarity (SSIM). SSIM is a metric that evaluates the similarity between images based on three components: luminance, contrast, and structural information, and it is known to correlate well with human visual perception. The SSIM between two images x and y is defined as follows:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where μ_x and μ_y are the means, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance of the two images. Constants C_1 and C_2 are used to stabilize the computation and avoid division by zero.

$$\text{Decision} = \begin{cases} \text{True} & \text{if } \mathrm{SSIM}(x,y) \geq T_1, \\ \text{False} & \text{otherwise} \end{cases}$$

If the SSIM value exceeds the threshold T_1, a random patch is cut from one image and pasted into the corresponding location of the other image. If the value is below the threshold, a new pair of images is selected. This process enables the automatic generation of pseudo-anomalous images that closely resemble real-world, subtle anomalies.

However, in some cases, the randomly selected patch may include only the background region, resulting in pseudo-anomalous images where the target object remains unchanged. To address this issue, as shown in Fig. 1(b), the similarity between the generated image and the original image is computed again.

$$\text{Decision} = \begin{cases} \text{False} & \text{if SSIM}(x, y) \geq T_2, \\ \text{True} & \text{otherwise} \end{cases}$$

If the SSIM value exceeds the threshold T_2, the generated image is considered to involve only background modifications and is discarded. Otherwise, it is added to the dataset as an anomalous image.

Using the dataset composed of normal and pseudo-anomalous images generated by the proposed method, we frame the task as a binary classification problem. For the model, we adopt a ResNet-18 pre-trained on ImageNet, with an additional two-layer MLP appended. For each object class, we generate 3000 pseudo-anomalous images for training.

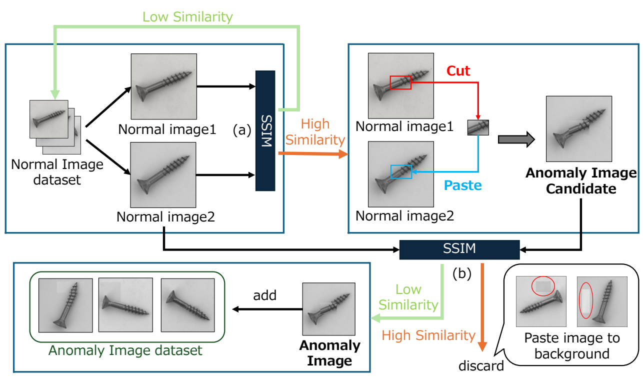

Fig. 1. Flow of the proposed method

III. Experiment

To verify the effectiveness of the proposed method, we conducted evaluation experiments. We compared the performance of an anomaly detection model trained on anomalous images generated by the conventional CutPaste method with that of a model trained on anomalous images generated by our proposed method.

We adopted the ROC Area Under the Curve (ROCAUC) as the evaluation metric. In our experiments, we used the MVTec AD dataset [3] and evaluated the performance across ten diverse object categories.

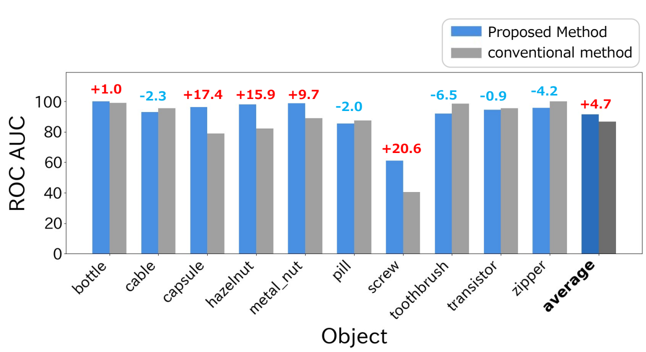

Fig. 2. Accuracy comparison for each object

The results shown in Fig. 2 confirm that the proposed method outperformed the conventional method in half of the object categories. Notably, for the screw category, the ROCAUC score improved by approximately 20%, while improvements of over 15% were observed for the capsule and hazelnut categories. Overall, the proposed method achieved an average accuracy improvement of 4.7% across the ten object categories.

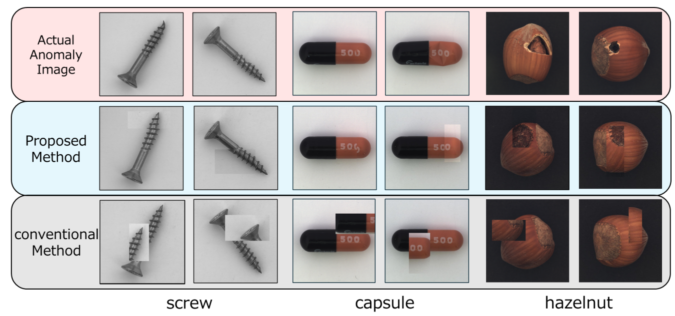

Fig. 3. Objects with significant accuracy improvement

Fig.3 shows examples of actual anomalous images, as well as those generated by the proposed and conventional methods, for the objects where accuracy improved significantly. We observed that in cases where conventional CutPaste generated visually unnatural anomalies, the proposed method achieved higher performance. For the screw category, which has large background regions and varying orientations among images, the proposed method was particularly effective. Additionally, compared to the conventional method, the proposed method generated images that were more similar to real-world anomalies, which likely contributed to the observed improvement in performance. For most objects, including screw, capsule, and hazelnut, the actual anomalies are small in terms of contour deviation. By reproducing such subtle anomalies, the proposed method achieved better performance than conventional CutPaste, which places patches at random positions.

IV. Conclusion

In this study, we proposed a CutPaste-based method for generating anomalous images with subtle pseudo anomalies. Compared to the conventional CutPaste approach, the proposed method generates more realistic anomalous images, contributing to improved performance in anomaly detection models. As future work, we aim to further enhance accuracy by incorporating self-supervised learning techniques.

References

[1] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, Tomas Pfister,“CutPaste: Self-Supervised Learning for Anomaly Detection and Localization”, *conference on computer vision and pattern recognition*, pp. 9664-9674, 2021.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, *conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[3] Paul Bergmann, Michael Fauser, David Sattlegger, Carsten Steger, “Mvtec AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection”, *conference on computer vision and pattern recognition*, pp. 9592-9600, 2019.