Study on Anomaly Classifier with Domain Adaptation

Chien Hung Wu, Rung Shiang Cheng, and Chi Han Chen(✉)

Overseas Chinese University, Taichung City, Taiwan
{ludwig1017,rscheng}@ocu.edu.tw

Abstract. There are various characteristics of industrial defects, and there is no fixed pattern to search for. Typically, anomaly detection models are used to identify defects. However, after experiencing a domain gap, industrial defect images often lead to a decrease in the verification accuracy of the source model. We conducted experiments to validate this and employed a domain adaptation model. Using color transformation algorithms, we generated source images with domain gaps and introduced them to a pre-trained model. We trained the model to learn features from the source domain and utilized a domain discriminator to differentiate between features from the source and target domains, assuming that the mappings of the target and source domains come from the same distribution. Comparative experimental results demonstrate that the domain adaptation model has a significant impact on improving accuracy. Specifically, the accuracy of the original "flower" category increased from 43.98% to 89.23%, and the "cable" category improved from 75.33% to 85.66%.

Keywords: manufacturing defect · anomaly detection · domain shift · domain adaptation

1 Introduction

In the industrial sector, many production lines generate numerous defective products during the manufacturing process. Consequently, there are various models aimed at addressing this issue. These models fall into three primary directions: reconstruction-based methods [1, 2], synthesis-based methods [3, 4], and embedding-based methods [5, 6]. These trends have emerged to tackle the challenge of industrial defect detection. However, these three methods often place excessive emphasis on the authenticity of synthesis, which can lead to misjudgments due to model generalization after image reconstruction. The embedding approach, which involves substantial computation, frequently consumes excessive system resources, impeding the real-time applicability of models in industrial settings. In practical industrial environments, there are distinct production lines, yet products of the same category need to be distinguished as defective or not. Different machines introduce variations in lighting and color space transformations, making it necessary to identify corresponding defects in images of defective products.

D.-J. Deng and J.-C. Chen (Eds.): SGIoT 2023, LNICST 557, pp. 3–11, 2024.
https://doi.org/10.1007/978-3-031-55976-1_1

If the above-mentioned methods encounter domain shifts under various conditions in practical industrial settings [1–6], these three directions are prone to increasing the model's misjudgment rate [1–6]. Consequently, traditional defect detection methods are incapable of accurately identifying defective products.

Regarding the proposition made in [7], when dealing with differences in color spaces, the model exhibits a noticeable decrease in accuracy. This is an unreasonable outcome considering the expected results. Through data augmentation techniques like adjusting color and brightness, the overall performance of the model should not decline. During training, a model's capabilities are limited to what it has learned. Consequently, the model may lack the ability to clearly recognize features that it has not been trained on. Therefore, even when learning images within the same color space, merely adjusting brightness during validation can lead to a decrease in model accuracy [7]. In the industrial domain, domain shifts are quite common, especially when images are captured using different cameras and lighting sources. Images of defects from various machines within the same production line introduce a wide range of domain shifts. This severely restricts the model's ability to generalize and can even necessitate training a separate model for each source domain with different lighting conditions. When models need to adapt to different environments, it can result in misjudgments. Images transformed through changes in color space retain the original image's contour features. Thus, theoretically, this should not lead to a decrease in model accuracy. However, after training the model with source images and validating them with domain shift caused by changes in color features due to image algorithms, a noticeable decrease in accuracy is observed (Table 1).

In this work, we begin with the industrial defect dataset collected from MVTEC AD. Initially, we employ image color transformation algorithms to convert the original RGB target images into grayscale 1-channel source images. The transformed images introduce a domain gap in the source data, and the objective is to assess whether the accuracy of defect detection in the data remains intact after this domain gap is introduced. Following this assessment, in cases where a decrease in accuracy is observed, we explore methods to maintain the original level of accuracy. Leveraging existing open-source Domain Adaptation models, we design an approach that specifically addresses domain gaps related to color features in images. It's worth noting that our approach focuses solely on

Table 1. The verification accuracy of the domain gap images compared to the source image reveals a noticeable decrease in precision. The model, when validated with source data on domain gap images, experiences a significant drop in accuracy. The model is unable to use a single weight to distinguish between features from different domains.

Type	Source_data	Domain_gap_data
Model	Resnet-18	Resnet-18
Flower	70.83%	43.98%
Cable	93.24%	75.33%
Hazelnut	95.00%	86.04%
Pill	43.02%	67.53%

color space transformations and does not involve algorithms such as image binarization or image transparency. Through our methodology, models trained using this approach can be applied to a broader range of images that experience shifts in color space due to domain shifts.

2 Related Works

2.1 Anomaly Detection

[8] This paper introduces a memory-based segmentation network (MemSeg) for defect detection and localization on the surfaces of industrial products within a semi-supervised framework. MemSeg utilizes a U-Net as its foundational architecture and introduces artificially generated anomaly samples and memory samples from both distinct and common perspectives to assist the network's learning process. MemSeg also incorporates a multi-scale feature fusion module and a spatial attention module to more effectively coordinate memory information and high-level features from input images. MemSeg achieves state-of-the-art performance on the MVTec AD dataset, with image-level and pixel-level AUC scores of 99.56% and 98.84%, respectively. This paper addresses the issue of distribution mismatch between the source domain and target domain, which can arise when using a pre-trained model to extract image features. Therefore, this paper employs a custom anomaly simulation strategy to enhance the model's adaptability to anomaly detection tasks.

[9] Introduces a simple Convolutional Neural Network (CNN) architecture called SimpleNet for the detection and localization of abnormal regions in images in an unsupervised manner. SimpleNet comprises four components: (1) a pre-trained feature extractor responsible for generating local features, (2) a shallow feature adapter used to transform local features into the target domain, (3) a straightforward abnormal feature generator, employed to simulate abnormal features by adding Gaussian noise to normal features, and (4) a binary abnormality discriminator used to differentiate abnormal features from normal ones. During the inference phase, the abnormal feature generator is discarded. SimpleNet achieves state-of-the-art performance on the MVTec AD dataset, with image-level and pixel-level AUC scores of 99.6% and 98.1%, respectively. The paper mentions the issue of inconsistency between pre-trained features and the target domain, thus employing a feature adapter to reduce domain bias and enhance the model's adaptability to the target domain.

2.2 Domain Adaptation

[10] Used for unsupervised domain adaptation in deep neural networks. Domain adaptation refers to the process of training an effective classifier or predictor when the data distribution in the training and testing sets differs. The authors' approach involves adding a domain classifier to the standard forward network, which is responsible for distinguishing features from the source domain and the target domain. Then, through the backpropagation algorithm, they simultaneously minimize the label prediction error for the source domain and maximize the error of the domain classifier. This process makes

the features invariant to domain shifts. The authors also introduce a gradient reversal layer to facilitate this optimization. This method can be applied to any deep network trained using backpropagation and has demonstrated exceptional performance in various image classification experiments, surpassing previous unsupervised domain adaptation methods. However, this paper does not consider other types of domain shifts, such as variations in label distribution, task objectives, or data patterns.

[11] Introduces an unsupervised domain adaptation method based on adversarial learning called Adversarial Discriminative Domain Adaptation (ADDA). This approach allows for learning a feature representation of the target domain without requiring target domain labels, enabling effective classification of target domain data. Initially, a discriminative model is trained on the labeled source domain to learn a feature mapping and a classifier for the source domain. Subsequently, through adversarial learning, a feature mapping for the target domain is learned in such a way that it can deceive a domain discriminator, making it unable to distinguish between source and target domain features. Finally, during testing, target domain data is mapped to a shared feature space and classified using the source domain classifier. This method presents a unified framework that generalizes existing adversarial domain adaptation methods as different design choices and demonstrates its superiority through experiments. The approach outperforms existing methods in standard cross-domain digit classification tasks as well as a more challenging cross-modal object classification task.

3 Methods

(See Fig. 1).

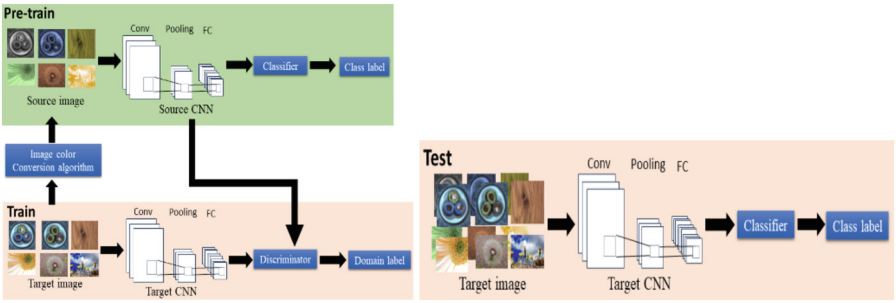

Fig. 1. Domain Adaptation model system architecture diagram

3.1 Generating Domain Gap Images

This paper focuses on the well-known MVTEC dataset in the field of anomaly detection and localization. An anomaly detection method is employed through a classification approach. Source images are generated using a color transformation algorithm applied to target images, which contain RGB color channels. No image binarization or transparency

adjustment is conducted. The RGB chromaticity of the original images is transformed algorithmically to match the source images, which are then used to train a pre-trained model. Feature generation from both the source and target images can be referenced (Fig. 2), revealing a noticeable domain gap between the two, as illustrated in the figures.

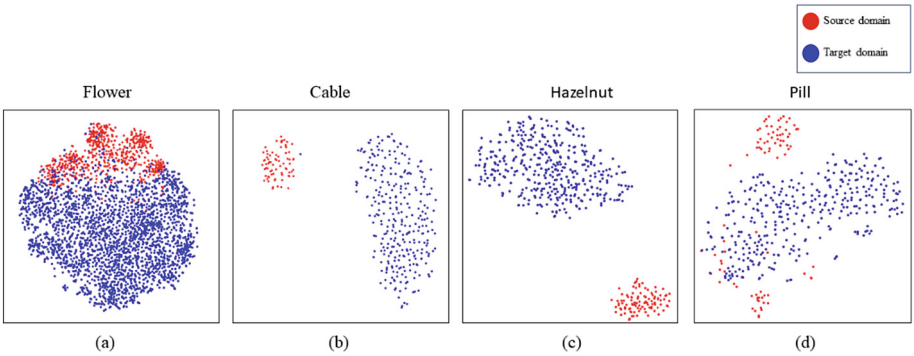

Fig. 2. The TSNE plot generated by the pre-trained model for (a) "Flower" (b) "Cable" (c) "Hazelnut" (d) "Pill" shows that the categories in the source and target do not overlap but gradually separate into distinct clusters. This plot serves as evidence that using a color transformation algorithm can indeed produce domain shifts.

3.2 Domain Adaptation Model

This paper verifies potential issues and a suitable solution for defect detection through a Domain Adaptation model. It begins by training a source model and processing target images through a color transformation algorithm to generate source images. The target images do not have labels. In the training phase, a pre-trained model is utilized to calculate classification losses, which are used to train the features and classifiers of the source domain. This ensures that they can correctly predict labels in the source domain. During training, two adversarial losses are employed. The first loss is computed by a discriminator to train a domain discriminator, enabling it to distinguish between feature mappings from the source and target domains. The second loss is used to train the target domain mapping, making it deceive the domain discriminator into thinking that mappings from the target domain and source domain originate from the same distribution. The objective is to align the feature spaces of the source and target domains. Once the feature spaces of the source and target domains are aligned, the model's recognition capabilities regarding images captured from different angles and color spaces are enhanced. This feature space alignment allows the model to adapt to various imaging scenarios.

3.3 Anomaly Detection with Classifier

Compared to performing segmentation for specific defects, using a classifier for defect detection offers greater versatility across various scenarios. Classifying the presence or

absence of defects, as opposed to traditional anomaly detection, provides more flexibility in terms of applications. Through this experiment's design and verification, it is demonstrated that models trained on classifiers and through domain adaptation methods can enable the adaptation of the model to a wider range of scenarios and lighting conditions.

4 Result

Through the Domain Adaptation model we used, the problems that may arise in defect detection and suitable solutions can be observed in various datasets. By applying our method to the same dataset, the accuracy of the original source domain can be maintained (Table. 2). By examining the TSNE plots, it is evident that there are distinct boundaries separating the features of the source and target within the Flower group (Fig. 3. a) and the Cable group (Fig. 3. b), Hazelnut grop (Fig. 3. c), Pill grop(Fig. 3. d). This indicates differences in feature distribution between the source and target domains, signifying that the model cannot effectively discriminate between different domain features using a single weight. After training, the Flower group shows improved classification accuracy in the target domain, gradually achieving more accurate classification between classes (Fig. 4. a). Additionally, after training, the feature distributions become more similar, without a significant domain gap observed as in Fig. 2 and Fig. 4. However, in the case of cable, where color distinction is crucial for defect detection, the accuracy is maintained, but there is no clear classification in the TSNE plot (Fig. 4. b) (Fig. 4. c) (Fig. 4. d).

Table 2. Comparing domain-shifted images with source images to validate the accuracy in comparison to our method, using the same dataset. From the data that originally reduced accuracy by generating domain-shifted images, our method effectively maintains accuracy without decreasing it. Furthermore, after the domain shift, it can effectively enhance accuracy.

Type	Source_data	Domain_gap_data
Model	Resnet18	Resnet18
Flower	70.83%	43.98%
Cable	93.24%	75.33%
Hazelnut	95.00%	86.04%
Pill	43.02%	67.53%
Model	Ours	Ours
Flower	72.93%	89.23%
Cable	94.59%	85.66%
Hazelnut	90.00%	86.28%
Pill	63.95%	68.97%

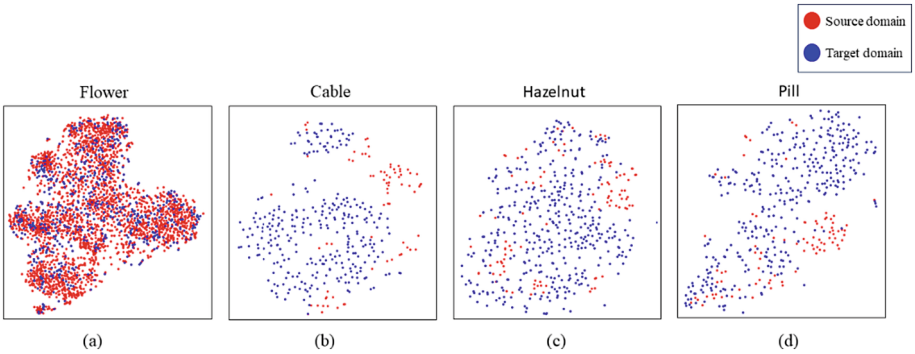

Fig. 3. Flower_Feature_TSNE(a)、Cable_Feature_TSNE(b)、Hazelnut_Feature_TSNE(c)、Pill_Feature_TSNE(d) (Red circles represent the source domain, and blue circles represent the target domain). It can be observed through the red circles of the source domain and the blue circles of the target domain that there is no clear boundary between the source and target. After training, the features of the source and target gradually converge.

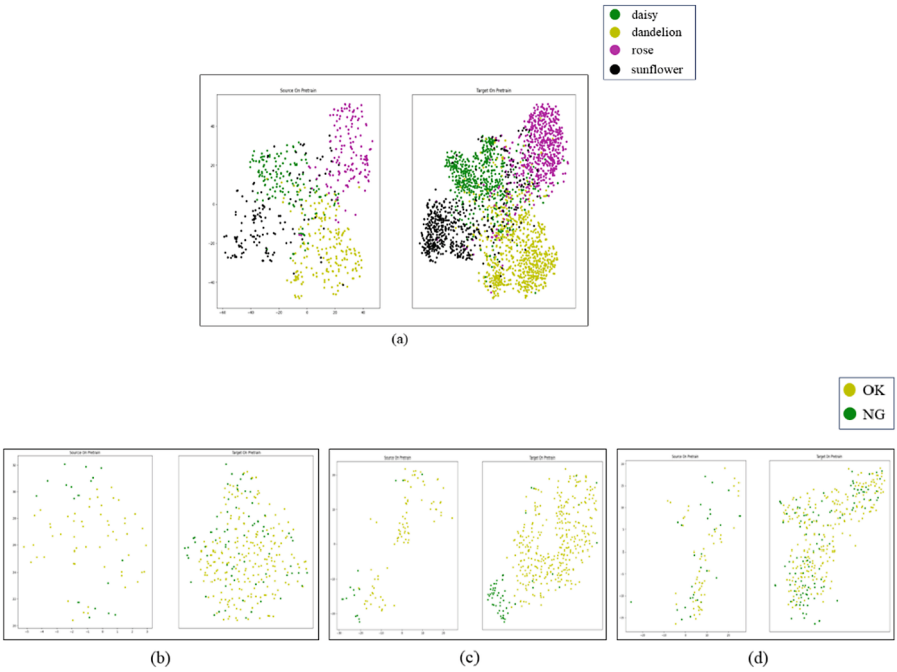

Fig. 4. Flower_classification_TSNE(a)、Cable_Classification_TSNE(b)、Hazelnut_Classification_TSNE(c)、Pill_Classification_TSNE(d) Through the TSNE plot based on classification, it can be seen that the features between the source and target gradually differentiate when compared for classification. Moreover, in terms of features (Fig. 2), they gradually converge, indicating a reduction in the domain gap issue.

5 Conclusions

According to the experimental results presented in this paper, when it comes to training and validating domain adaptation for generating images with domain gaps in the defect detection domain, noticeable differences in accuracy can be observed across different domains. This is due to the fact that industrial production lines often involve various lighting sources and image capture methods. Therefore, domain gaps in captured images are likely to occur. Through the experimental methodology we have designed, simulating domain gaps in images between the source and target domains allows us to maintain the classifier's accuracy in defect classification. In the future, this approach can be applied to various production lines where different domains of images need to be trained. By training the model with images from different domains, we can enhance the model's ability to generalize across diverse scenarios.

Acknowledgment. The authors would like to thank the National Science and Technology Council, Taiwan, R. O. C. for financially supporting this research under Contract No. NSTC 112-2221-E-240-002-MY3.

References

1. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., van den Hengel, A.: Memorizing normality to detect anomaly: Memory-augmented deep auto encoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1705–1714 (2019)
2. Ristea, N.-C., Madan, N., Ionescu, R.T., Nasrollahi, K., Shahbaz Khan, F., Moeslund, T.B., Shah, M.: Self-supervised predictive convolutional attentive block for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13576–13586 (2022)
3. Li, C.-L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9664–9674 (2021)
4. Zavrtanik, V., Kristan, M., Skocaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8330–8339 (2021)
5. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: Del Bimbo, A., et al. (eds.) Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV, pp. 475–489. Springer International Publishing, Cham (2021)
6. Roth, K., Pemula, L., Zepeda, J., Scholkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14318–14328 (2022)
7. De, K., Pedersen, M.: Impact of colour on robustness of deep neural networks. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 21–30. https://doi.org/10.1109/ICCVW54120.2021.00009
8. Yang, M., Wu, P., Feng, H.: MemSeg: a semi-supervised method for image surface defect detection using differences and commonalities. Eng. Appl. Artif. Intell. **119**, 105835 (2023)

9. Liu, Z., Li, C.-L., Sohn, K., Yoon, J., Pfister, T.: SimpleNet: a simple network for image anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
10. Ganin, Y., Lempitsky, V.: Unsupervised Domain Adaptation by Backpropagation. In: International Conference on Machine Learning (ICML), pp. 1180–1189 (2015)
11. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Computer Vision and Pattern Recognition (CVPR) (2017)