Simultaneous Optimization of Abnormality Discriminator and Illumination Conditions for Image Inspection of Textile Products

Yuma Nishikawa, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Japan
{*sakaue, junsato*}*@nitech.ac.jp*

Keywords: Anomaly Inspection, Textile Product Inspection, Illumination Optimization.

Abstract: In this study, we propose a method to simultaneously learn and optimize the illumination conditions suitable for anomaly inspection and a neural network for anomaly inspection in textile product anomaly inspection. In the inspection of abnormalities in industrial products such as textile products, it is necessary to optimize the imaging environment including the lighting environment, but this process is mostly done manually by trial and error. In this study, we show that highly accurate inspection of abnormalities can be achieved by using a display whose light source position and brightness can be easily changed, and by presenting a proof pattern suitable for abnormalities on the display. Furthermore, we show how to simultaneously optimize the neural network and illumination conditions used for such anomaly inspection. We also show that the proposed method can appropriately detect anomalies using images actually taken.

1 INTRODUCTION

The process of product inspection at industrial product manufacturing operations is essential to assure the quality of manufactured products. In this inspection process, products are checked for any abnormalities, such as scratches and irregularities in coloring, which can occur in a wide range of products. Therefore, it is difficult to automatically determine abnormalities using image processing technology, etc., even if they are easily determined to be abnormal by humans. Therefore, methods to automatically determine abnormalities are still being researched and developed. These methods can be classified into those that learn only normal data and detect abnormalities[(Akcay et al., 2018; Akçay et al., 2019)], and those that collect both abnormal and normal data and learn the differences between them[(Bergmann et al., 2019; Bergman et al., 2020)]. Since each of these methods can be used in different situations, they are used for different purposes.

It is well known that the inspection of abnormalities in textile products such as fabrics is particularly difficult. This is partly due to the fact that textile products have various patterns and colors, and the appearance of textile products varies greatly depending on these patterns and colors when inspecting for abnormalities. In addition, textile products are composed of fine threads, so that even a slight change in the condition of the fibers or the surrounding conditions can change the appearance of the textile product. This means that the appearance of anomalies in textile products also changes significantly depending on the conditions under which the textile product is imaged. Figure 1 is an example of an image of the same flaw (abnormality) taken under different light source environments, and it can be confirmed that the appearance changes depending on the situation. Due to these characteristics, human inspectors visually inspect by changing the direction of observation and illumination conditions to create a situation in which it is easy to detect anomalies.

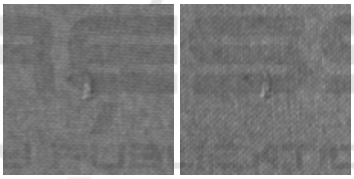


Figure 1: Example of change in appearance of flaws on textile products.

Considering the above, it is necessary to optimize the arrangement of cameras, illumination, etc. for abnormality inspection in accordance with the abnormality to be detected. The importance of optimizing the imaging environment is well known in image processing-based anomaly inspection, and when constructing an actual system, the optimal imaging con-

Nishikawa, Y., Sakaue, F. and Sato, J.
Simultaneous Optimization of Abnormality Discriminator and Illumination Conditions for Image Inspection of Textile Products.
DOI: 10.5220/0013322300003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 755-760
ISBN: 978-989-758-728-3; ISSN: 2184-4321

ditions are studied by repeatedly capturing images of actual products. Since this process requires a lot of time and cost, an efficient method to perform it is needed. In this study, we propose a method for automatically optimizing the imaging conditions.

For this purpose, we propose a method to perform highly accurate anomaly inspection by using a display used for video presentation as a light source to illuminate the object and optimizing the illumination patterns displayed on the display. There are an almost infinite number of patterns that can be displayed on a display used as a light source, and it is possible to create a very complex illumination environment. On the other hand, the number of variations makes it impractical to optimize them through trial-and-error. Therefore, the proposed method, which can automatically optimize the lighting environment, has a great advantage when constructing an abnormality inspection system. In addition, the proposed method trains a neural network used for anomaly detection in conjunction with such lighting pattern optimization. Combined with the derived patterns, the proposed method can detect anomalies with extremely high accuracy.

2 ANOMALY DISCRIMINATION USING VISION TRANSFORMER (ViT)

We first describe methods for detecting anomalies using image information. Two methods are possible for image abnormality detection: one is to use only normal images and discriminate deviations from them as abnormalities(Perera and Patel, 2019; Bergmann et al., 2019), and the other is to collect abnormal and normal images in advance and discriminate them by 2-class identification. While the former method is easy to collect training data, it has the problem that it is difficult to improve the accuracy of abnormality discrimination compared to the method that directly uses abnormality data. In this paper, assuming that the anomaly images can be collected to some extent, we proceed with the discussion with the main objective of anomaly detection by 2-class discrimination.

In traditional anomaly detection before deep learning, classification of abnormal and normal images has been attempted based on the statistical properties of image features, such as principal component analysis and linear discriminant analysis. In these discriminant analyses, various types of information have been obtained not only by directly using image intensity information, but also by using information such as edges obtained by filtering. On the other hand, in recent years, methods using neural networks represented by CNNs (Convolutional Neural Networks) have been widely used in anomaly detection(Perera and Patel, 2019). Recently, neural network structures that do not use convolution, called Vision Transformer (ViT) in particular, have been widely used,(Dosovitskiy et al., 2021). Transformer architecture in neural networks is a method proposed in the field of natural language processing, and the structure that applies this to image information is the Vision Transformer, which is widely used for image identification.

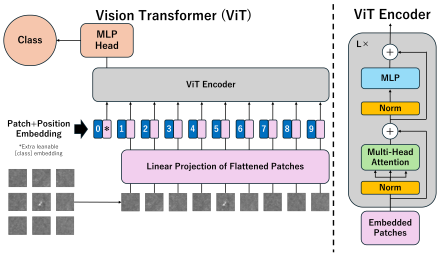


Figure 2: Representative architecture of Vision Transfomer.

The Vision Transformer divides the input image into a set of small patches and processes these as tokens using the Transformer structure. In this case, each token is input to a network structure called ViT Encoder, as shown in Fig.2. The ViT Encoder uses a structure called “multi-head attention” to represent the relationship between tokens, enabling it to capture global image features even from a finely segmented patch set. The features obtained by the Encoder are input to MLP (Multi-Layer Perceptron), which calculates the final identification result.

It is known that a large amount of training data is required to achieve sufficient performance when constructing a discriminator using Vision Transformer, compared to using a CNN. However, it is known that this problem can be avoided by training (pre-training) a neural network in advance using a large amount of image data, and then learning the neural network in transfer learning according to the task. This has shown that the Vision Transformer can be applied to a variety of tasks even when training data is limited(Dosovitskiy et al., 2021). In the discriminator used in this study, we aim to achieve highly accurate anomaly discrimination by transfer learning a neural network that has been trained with a large amount of data in advance, using the target data.

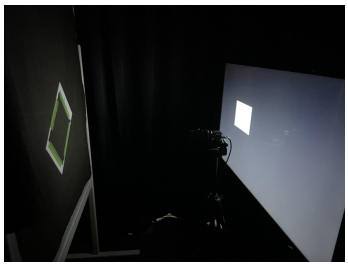


Figure 3: Imaging environment with display as light source.

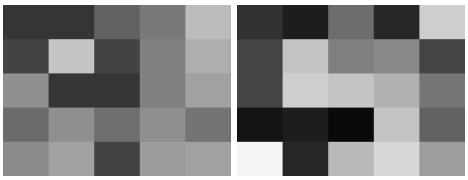


Figure 4: Example of a pattern on a display.

3 SIMULTANEOUS OPTIMIZATION OF ILLUMINATION CONDITIONS AND ANOMALY DETECTION ViT

3.1 Lighting Environment Using Controllable Lighting

In this study, anomaly discrimination is performed using ViT as described in the previous section. However, in order to achieve stable and robust abnormality detection, it is necessary to acquire optimal images suitable for abnormality detection. As mentioned above, the appearance of a textile product, which is composed of a collection of fine threads, varies greatly depending on the lighting environment in which it is photographed. Therefore, images taken under inappropriate illumination conditions do not sufficiently show the characteristics of the textile in the image, making it difficult to perform proper abnormality detection. To avoid this problem, a method of capturing images of objects under various illumination conditions in advance and using these images for identification can be considered. However, this method increases the time required to capture images and increases the cost of equipment. Therefore, it is necessary to prepare optimal illumination conditions suitable for abnormality detection in order to perform abnormality identification at the lowest possible cost and in the shortest possible imaging time.

3.2 Determination of Illumination Environment Using 1×1 Convolution

When a display is used as a light source, there is a nearly infinite variation of possible illumination patterns. Therefore, it is not realistic to optimize them manually. In this study, we propose a method to simultaneously optimize the illumination conditions for anomaly detection and the neural network for anomaly discrimination by utilizing the linearity of the light source in the image.

In this method, a rectangle with a certain brightness is displayed on the display as shown in Fig.3 and is used as the light source. These rectangles divide the display area and reproduce various lighting environments by changing the brightness of each rectangle as shown in Fig.4. Let N be the number of light sources (rectangles) displayed on the display, and let $\mathbf{I}_i$ be the image captured in an environment where only each light source is turned on. In this case, the image $\mathbf{I}$ captured by lighting multiple light sources simultaneously can be expressed by the following equation.

In this study, a large display is used as illumination, as shown in Fig.3, and the aim is to realize highly accurate anomaly discrimination by controlling it appropriately. The display used here can emit light at arbitrary locations and at arbitrary brightness. By displaying various patterns on the display, a very complex lighting environment can be created. Therefore, by estimating and presenting display patterns suitable for anomaly detection, it is expected to be possible to create a situation in which anomaly detection can be performed with higher accuracy than with general lighting.

$$\mathbf{I} = \sum_{i=1}^{N} E_i \mathbf{I}_i, \tag{1}$$

where E_i is the brightness of each light source and is the ratio of the brightness when $\mathbf{I}_i$ is taken. When $\mathbf{I}_i$ is illuminated at the maximum brightness that can be displayed when $\mathbf{I}_i$ is taken, $0 \leq E_i \leq 1$. Therefore, the determination of the optimal lighting environment can be said to be the determination of E_i as shown in the Eq.(1).

In determining E_i, we consider the simultaneous optimization of these parameters and the neural network for anomaly detection. As shown in Eq. (1), an image taken under optimal illumination conditions can be expressed as a weighted sum of images taken under a single light source. If N images are considered as N channel images, the estimation of the optimal illumination image can be thought of as the process of merging N channel images into a single

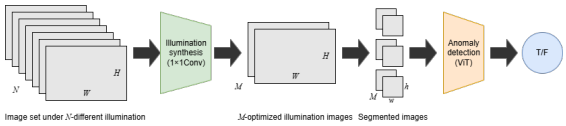


Figure 5: Proposed architecture.

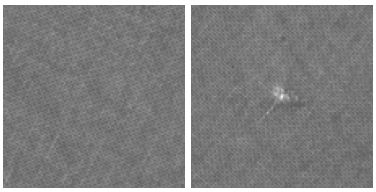
channel image. Since the linear sum of images is synonymous with the linear sum of each pixel in the image, the linear sum of images can be expressed as the convolution of $1 \times 1 \times N$ kernels(Ueda et al., 2024). Therefore, the determination of optimal illumination conditions can be redefined as the optimization of this convolution kernel. Based on the above, we define the network structure including the lighting optimization as shown in Fig.5

In Fig.5, N images are input and M images are obtained by the convolution of N images with M kinds of 1×1 kernels. By inputting this to the ViT architecture, it determines whether the input image is a normal image or an abnormal image. This convolution process with 1×1 kernel is equivalent to the process described in Eq(1). Therefore, the image obtained by this convolution is equal to the image taken under the illumination pattern based on the kernel weights. Therefore, while N images taken under each light source must be acquired when training the network, only M images taken under the obtained optimal illumination can be used for inspection. Since M can be any integer greater than or equal to 1, it can be set according to the allowable imaging time and required inspection accuracy to accommodate various situations. This allows the neural network to learn and the illumination conditions suitable for it at the same time.

4 EXPERIMENTAL RESULTS

4.1 Environment

In this section, we present the results of an experiment conducted to verify the effectiveness of the proposed method in the previous section. In this experiment, the proposed method was used to simultaneously train the illumination condition and the neural network, which was then used to detect anomalies on the fabric product. For this purpose, we took images of flaws on the fabric using a needle, and examined whether it is possible to discriminate between flawless areas (normal areas) and flawed areas (abnormal areas). Examples of normal and abnormal images are shown in Fig.6 These images were captured under the illumination of the display shown in Fig.3. The input images are the 224×224 images extracted from the captured images. The Vision Transformer used to discriminate anomalies is a pretrained one using JFT-300M(Sun et al., 2017) and ImageNet1k(Russakovsky et al., 2015). This model was also trained as a 2-class discriminator with a 1×1 convolutional kernel that indicates the illumination condition using the structure shown in Fig.5. Since ViT is trained with RGB 3-channel images as input, after reproducing the illumination condition using the 1×1 kernel, three 3×3 convolutional kernels were applied to convert the image to a 3-channel image, which was used as input to the discriminator.



(a) Normal image (b) Abnormal image

Figure 6: Examples of input image.

On the display, a rectangle is displayed as shown in Fig.3, which is used as the light source. Twenty-five (5×5) are displayed on the device. The brightness of each light source was determined by the proposed method. The number of kernels to be trained was set to 1 (one captured image) and 2 (two captured images), and the performance of each case was examined. For training, 50 sets of normal images and 50 sets of abnormal images were used, and 1000 sets of data (500 normal images and 500 abnormal images) were used for evaluation. For comparison, additional training of the discriminator was performed on images taken under a single illumination using the same data, and the performance of the discriminator was examined. For the comparison method, the lighting condition with the best discrimination result among 25 input images was selected as the first image, and for the second image, the condition with the best discrimination result was selected when discrimination was performed using two images, the lighting condition with the best discrimination result and another image.

The evaluation value was set to be rejected if it was less than a threshold value, and by changing this threshold value, the rate at which an abnormal image was judged as a normal image (false acceptance rate: FAR) and the rate at which a normal image was erroneously rejected (false rejection rate: FRR) were obtained, and each method was evaluated.

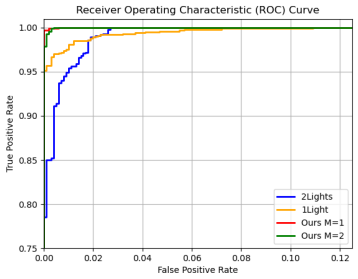
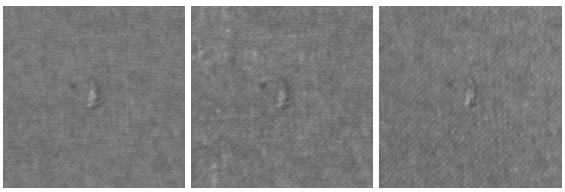


Figure 7: ROC curves for proposed and conventional ViT.



(a) Optimal illumination (b) Best single illumination (c) Worst single illumination

Figure 8: Images taken by the proposed method (a), images taken with the best single illumination source (b), and images taken with the worst single illumination source (c).

4.2 Results

First, the ROC curves of the proposed method and the method using a single illuminated image are shown in Fig.7. This figure shows that all methods have high discrimination performance, which confirms the effectiveness of ViT in discriminating anomalies. Among the methods, the proposed method shows the best discrimination performance when two input images are used, which confirms its effectiveness. Examples of images actually taken under the optimal illumination, images taken in the environment with the best discrimination performance, and images taken in the environment with the worst discrimination performance are shown in Fig.8. These images show that in the image taken under the optimal illumination, only the abnormal area is emphasized, making it easier to discriminate abnormalities. On the other hand, in the worst-case environment, it is difficult to see the abnormal area, and this is thought to make it difficult to discriminate abnormalities.

Next, the false rejection rate (FRR) when the false acceptance rate (FAR) is set to 0 is shown in Tab.1. This corresponds to over-detection, in which a normal part is judged as abnormal in the inspection process. Since missing anomalies are not acceptable in product inspection, it is particularly important from the standpoint of practicality how low the over-detection can be suppressed when the over-detection is set to zero. The table shows that the proposed method can suppress the false rejection rate to a low level even when the number of input images is small. This indicates that the proposed method can be used to construct a very practical anomaly detection system.

Table 1: False rejection rate (FRR) with false acceptance rate set to 0[%].

	FRR[%]
Proposed (1 image)	0.6
Proposed (2 images)	0.4
Single illumination (1 image)	10.9
Single illumination (2 images)	2.7

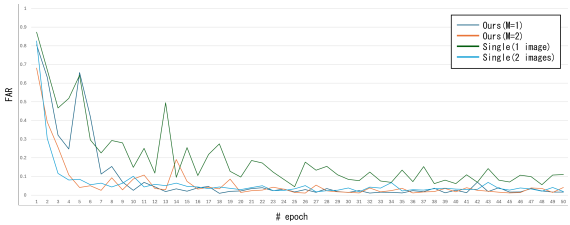


Figure 9: Change in false rejection rate with learning.

In addition, how the FAR changed with learning is shown in Fig.9. The results show that the proposed method with illumination optimization not only keeps the FAR low, but also always keeps the FRR low regardless of the learning process. This indicates that the proposed method is capable of stable learning. These results confirm that the method proposed in this paper can construct a stable and robust anomaly detection system.

5 CONCLUSION

In this study, we proposed a method to improve the accuracy of textile product abnormality inspection by simultaneously optimizing the illumination conditions and the abnormality detection neural network.

REFERENCES

Akcay, S., Atapour-Abarghouei, A., and Breckon, T. P. (2018). Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer.

Akçay, S., Atapour-Abarghouei, A., and Breckon, T. P. (2019). Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Bergman, L., Cohen, N., and Hoshen, Y. (2020). Deep nearest neighbor anomaly detection. *arXiv preprint arXiv:2002.10445*.

Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., and Steger, C. (2019). Improving unsupervised defect segmentation by applying structural similarity to autoencoders.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.

Perera, P. and Patel, V. M. (2019). Learning deep features for one-class classification. *IEEE Transactions on Image Processing*, 28(11):5450–5463.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proc. ICCV2017*.

Ueda, T., Kawahara, R., and Okabe, T. (2024). Learning projection patterns for direct-global separation. In *Proc. the 19th International Conference on Computer Vision Theory and Applications (VISAPP2024)*, pages 599–606.