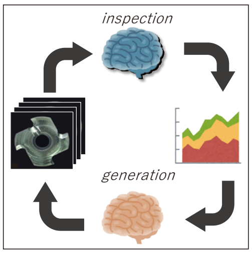


Accelerating the Development of Visual Inspection AI System Using Image Generation AI

Shun YOSHIDA*, Shoya ISHIDA, and Katsushi MIURA

Our company has traditionally relied on manual visual inspections for defect detection, which required considerable manpower and often led to inconsistencies in inspection standards depending on individual inspectors. To address these issues, we developed an AI-based visual inspection system to automate defect detection. However, a key challenge was the limited availability of defective product images for training, which impeded the enhancement of detection accuracy. To resolve this, we introduced an image generation AI capable of producing a large number of synthetic defective images from a small number of real defective samples. Furthermore, we developed a complementary method to analyze and identify the types of images that the inspection AI system is prone to misclassifying, thereby pinpointing its weaknesses. By generating targeted images to address these weaknesses and iteratively training the inspection AI system in a focused manner—a process we term the "anti-weakness training loop"—we were able to significantly improve system performance. Consequently, even with limited real defect data, we shortened the development time and enhanced the detection accuracy of the visual inspection AI system. This paper reports on the development and results of this approach.

Keywords: image recognition, visual inspection, deep learning, image generation AI

1. Introduction

Our company developed an AI-based visual inspection system that uses image recognition AI in order to automate the sensory evaluation and quantitative evaluation of image data in visual inspection and product development. This inspection system has already been used in various processes. This system, which uses deep learning,*1 requires a large amount of training data, generally a few thousand or more each for good and defective images to ensure high performance operation.(1) However, the ratio of defects found on manufacturing site is small, and a long period of time—from several months to several years—is required to acquire the defective images needed to train this system. This results in a longer development time for AI-based visual inspection systems.

To address this issue, we focused on image generation AI technology, which is attracting much attention in recent years. As the name implies, this AI is capable of generating images (synthetic images) which are very similar to actual images (real images). In addition, advanced image generation AI can control the appearance of the images it generates, as shown in Fig. 1,(2),(3) and can freely change the position, size, and color of defects on the images. This technology makes it possible to identify image patterns that the visual inspection system tends to misclassify—revealing its weakness. Furthermore, this technology can effectively generate a large amount of training data from only a few real images collected on the manufacturing site. This data is particularly effective in addressing the weaknesses of visual inspection systems, thereby accelerating the development of AI-based visual inspection.

This paper outlines anti-weakness training loop technology and its composing elements in Chapter 2, together with the application results of this technology in Chapter 3.

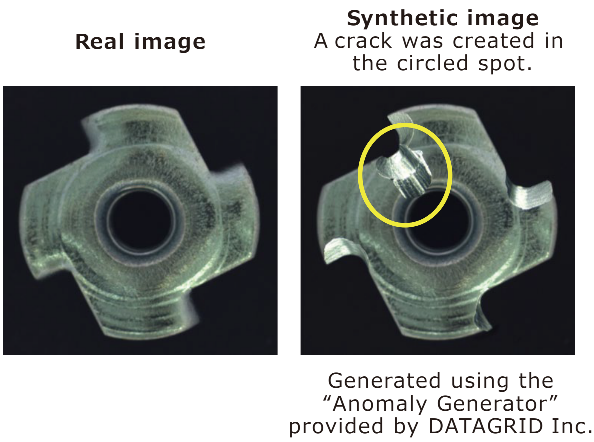


Fig. 1. Real image and synthetic image generated by image generation AI

2. Outline of Anti-Weakness Training Loop

Figure 2 shows the system for training the defect detection model for visual inspection systems. The system consists of four major steps as shown below.

Step 1: Acquisition of real images
Step 2: Generation of synthetic defective images
Step 3: Training and evaluation of defect detection model
Step 4: Analysis of misclassification results (Then, return to Step 2.)

After returning to Step 2, a large number of synthetic images are generated, focusing on the weaknesses that were not correctly classified in Step 4, and the defect detection model is retrained. In this paper, this process of repeat-

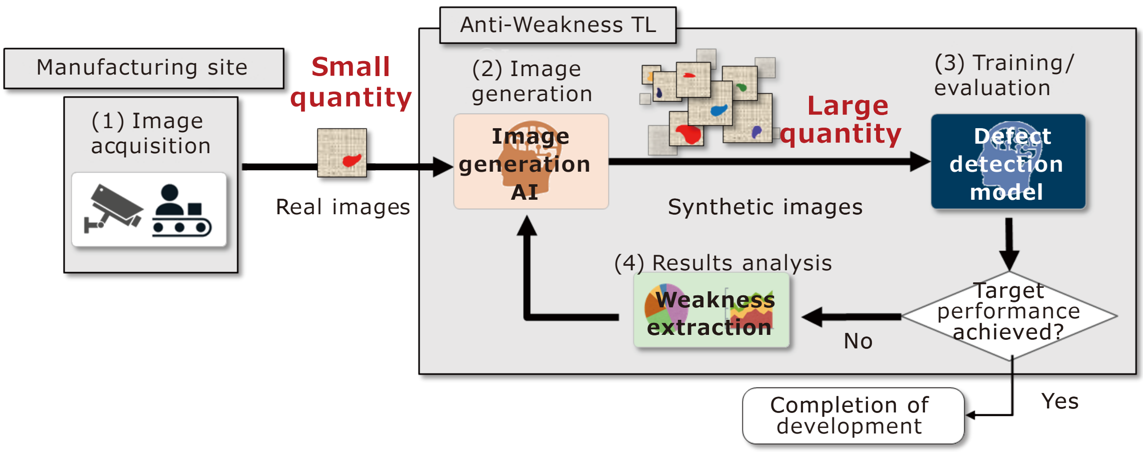


Fig. 2. Outline of anti-weakness training loop

edly discovering and training the weakness of the defect detection model is referred to as the "anti-weakness Training Loop (TL)" by making an analogy to the "training" of people. We caried out the development of this technology in collaboration with DATAGRID Inc.,*[2] a company having a technological advantage in the generation of synthetic images from a small number of images and providing the "Anomaly Generator," AI for generating defective product images. This technology is applicable to any task of image classification, object detection, and segmentation.*[3] In the following sections, each step of the anti-weakness TL is described in detail as a two-class classification task for goods and defects.

2-1 Image acquisition on manufacturing site

In Step 1, good and defective images of the target product are collected. There are no special requirements for the image acquisition process; a commercially available general-purpose camera can be used. Both color and grayscale images are acceptable. These images are utilized to train the image generation AI. While the minimum requirement is one good image and one defective image, it is empirically preferable to collect at least 10 images of each.

2-2 Generation of synthetic defective image

In Step 2, synthetic defective images used to train the defect detection model are generated. The synthetic image generation system employed in this study is shown in Fig. 3. Both a real good image and a defective image are provided as input to the image generation AI, together with instructions on how the appearance of the defect should be modified. By combining good and defective images in this way, the system can generate a large number of realistic synthetic defective images from as few as 10 real defective images. Training the defect detection model with a diverse set of synthetic images—varying in defect location, size, and color—improves the performance of defect detection model. In a validation experiment, experienced inspectors evaluated the synthetic images and judged them to be indistinguishable from real defects at a high rate of 81.1%.

The image generation AI enables the generation of an arbitrary number of synthetic images. Since the appearance of the defective region can be specified, it is possible to selectively generate rare defect modes that seldom occur on the manufacturing site, or to suppress the generation of unnatural defects that are known to be unlikely to occur in practice. Examples of the appearance of a defect that can be specified (called "appearance property") are shown in Table 1.

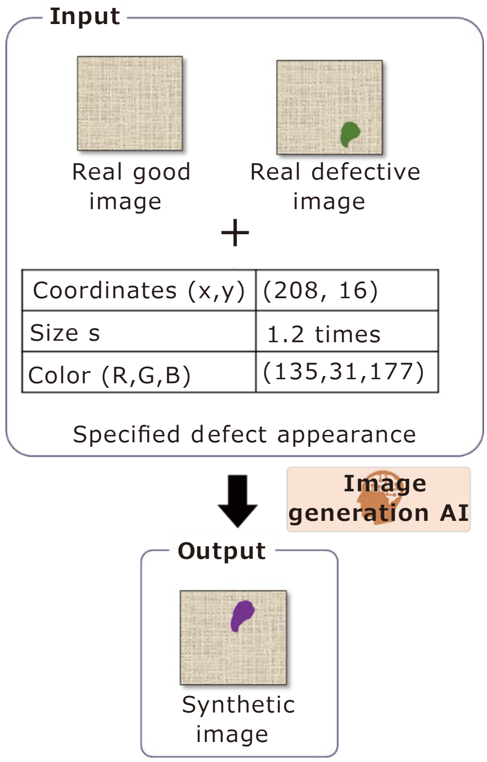


Fig. 3. Synthetic defective image generation system

Table 1. Examples of an appearance property that can be specified by image generation AI

Defect appearance	Description
Coordinate (x,y)	Location of the defect
Size s	Scaling ratio of the defect
Color (R,G,B)	Color of the defect

2-3 Training and evaluation of the defect detection model

In Step 3, the defect detection model is trained with both a small number of real images and a large number of synthetic defective images. Any image classification model can be used for this purpose; in particular, Vision Transformer[4] and ResNet[5] are widely recognized as effective models in the field of image classification.

After training of the defect detection model, its performance is evaluated using test data. If the evaluation results meet or exceed the target performance, the training process is considered complete. Otherwise, the procedure proceeds to Step 4.

2-4 Analysis of misclassification results

In Step 4, a detailed analysis is conducted to identify the image appearance that cause the defect detection model to often make misclassifications, i.e., the images that degrade its performance. The test data used for this analysis is also generated by image generation AI. However, if the same image generation AI is used to generate both the training data and test data, data leakage*4 will occur, thereby preventing a proper analysis of the results of the defect detection model. To avoid this problem, a separate image generation AI trained on a different image dataset is prepared for the evaluation, in addition to the image generation AI used in Step 2.

The analysis procedure is as follows. First, at least one appearance property is selected for evaluation, and test data are generated by varying the value of that property at equal intervals using the evaluation image generation AI. When analyzing the influence of detect location on model performance, multiple synthetic images are generated with equally spaced x- and y-coordinate values. These images are then used as test data (Fig. 4). In this manner, weaknesses of the defect detection model with respect to defect location can be visually identified by plotting the misclassification rate for each coordinate, as shown in Fig. 5. In this analysis, the appearance property consisted of two coordinates—x and y—allowing the accuracy of the model to be visualized as a two-dimensional heat map. Note that when at least four appearance properties are selected, it becomes difficult to visualize the model's weakness, as the accuracy can no longer be illustrated effectively.

2-5 From the second training loop onward

Referring to Fig. 2, the procedure from Step 4 back to Step 2 is described below. In Step 2 of the second and subsequent loops, the image generation AI trained during the first loop is used to generate synthetic images that reflects the appearance properties associated with the model's weaknesses. In Step 3, these newly generated images are added to the training data to retrain the defect detection model. After training, the performance of the model is evaluated using the same test data that were used in the first Step 3. If the evaluation results meet or exceed the target performance, the development of the defect detection model is considered complete. If the performance remains below the target, the model's weaknesses are re-analyzed in Step 4, and the procedure continues to the next training loop.

Since this training loop can be repeated multiple times, the performance of the defect detection model can be further improved through iteration.

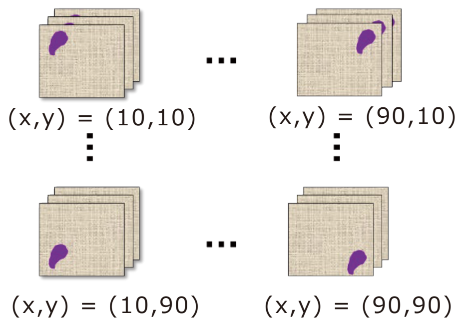


Fig. 4. Generation of evaluation test data

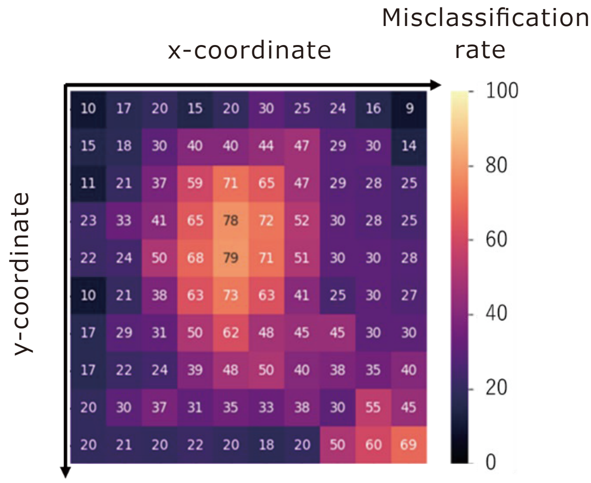


Fig. 5. Misclassification rate of the defect detection model at each coordinate
The higher misclassification rates in the brighter areas indicate a weakness of the model

The image generation AI only needs to be trained once at the beginning, and the test data used for weakness analysis in Step 4 can be reused in all subsequent loops. Thus, the training cost of the image generation AI and the cost of generating evaluation test data do not increase proportionally with the number of training loops.

3. Experimental Results

This chapter presents the results of applying the proposed anti-weakness TL to one of our products. The target was an image of a ceramic material containing a white-colored defect inside the yellow circle shown in Fig. 6. A two-class classification task was adopted to deter-

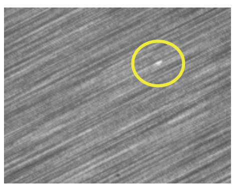


Fig. 6. Image of the ceramic material used in the experiment

mine whether the image was good or defective. The evaluation metric was the miss-detection rate—i.e., the rate at which defective images were incorrectly classified as good. Good images in the training data consisted of 500 real images, while the defective images were varied depending on the experimental conditions.

3-1 Weakness extraction of the defect detection model

This section describes the method for extracting weaknesses of the defect detection model from the distribution of the accuracy rates corresponding to each appearance property, obtained in Step 4 of the anti-weakness TL.

As Fig. 5 shows, the distribution of the miss-detection rate usually varies gradually, with a certain point exhibiting the highest rate. This suggests that images with similar appearances tend to yield similar miss-detection rates. A simple approach to identifying weaknesses is to regard all points where the miss-detection rates exceeding a predefined threshold as weaknesses. However, such high-rate points are often concentrated in localized regions of the appearance-property space, resulting in the selection of visually similar properties as weaknesses. Consequently, only synthetic images with highly similar appearance are added to the training data, which biases the diversity of training data. This may lead to degraded model performance or require additional training loops to achieve the target accuracy.

To avoid treating visually similar appearances repeatedly as weaknesses and to efficiently extract a diverse set of weaknesses, we chose to search only for the local maxima in the miss-detection rate distribution. When p appearance properties are selected, the miss-detection rate distribution can be represented in a p-dimensional space. This space is uniformly divided into small subregions of size k^p, where k is a tunable parameter that determines the granularity of the division and affects the performance of the defect detection model. For example, in the case of $p = 2$ and $k = 2$, the 2D space is divided into subregions as shown by the red lines in Fig. 7 (a). From each subregion, only one candidate weakness is extracted in order to avoid selecting visually similar weaknesses. Furthermore, any candidate whose miss-detection rate is lower than a predefined threshold r is excluded from consideration, as shown in Fig. 7 (b). This thresholding prevents an excessive number of weaknesses from being identified unnecessarily. A likely reason for the existence of weaknesses in the defect detection model is the insufficient presence of images with those appearance properties in the training dataset. To address this, synthetic images similar to the appearance properties of each weakness are generated in proportion to their miss-detection rates and added to the training data. This allows the defect detection model to learn from previously underrepresented appearances and is expected to improve its overall performance. Therefore, a "weakness score" is computed based on the miss-detection rate of each extracted weakness, as shown in Fig. 7 (c). This score serves as a guideline in Step 2 of the next training loop, determining how many synthetic images should be generated for each weakness.

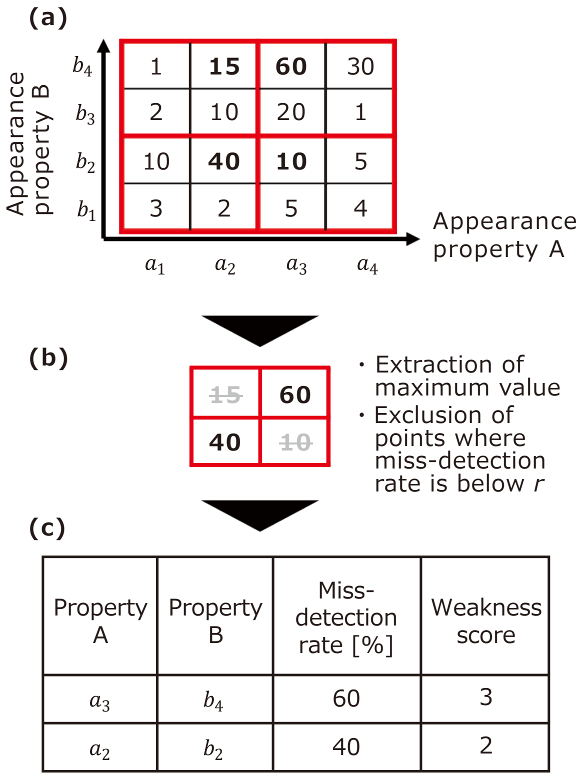


Fig. 7. Weakness extraction algorithm

3-2 Effectiveness of synthetic image

We evaluated the effectiveness of synthetic defective images generated by the image generation AI in reducing the miss-detection rate of the defect detection model. To augment the training data, 10 real defective images were increased to 1,000 images using commonly used augmentation techniques such as vertical and horizontal flipping. Two types of test data were prepared: (1) a real defect test data consisting of 140 real defective images and (2) a synthetic defect test data consisting of 1,000 synthetic defective images. The evaluation results are shown in Table 2. The miss-detection rate was lower when both real and synthetic defective images were used for training, compared to training with real images only. This indicates that synthetic defective images contributed to improving the defect detection model's performance. On the other hand, when the model was trained using only synthetic defective images, the miss-detection rate decreased on the synthetic test data but increased on the real test data. This degradation is likely due to the presence of artifacts or

Table 2. Change in miss-detection rate when synthetic images are used in training

Training data (Images [Count])		Evaluation results (Miss-detection rate [%])	
Real defect	Synthetic defect	Real defect	Synthetic defect
1000	0	23.8	11.4
1000	1000	**14.3**	7.2
0	1000	26.4	4.8

Note: Training with only synthetic defective images was conducted for experimental comparison and is not recommended for practical use.

noise patterns unique to synthetic images,[6] which are not present in real images. When trained sorely on synthetic images, the model may incorrectly learn such artifacts as features of defects, leading to degraded performance on real data. Therefore, it is effective to use both real and synthetic images as the training data.

3-3 Ratio of real to synthetic image

Based on the previous findings that both real and synthetic images are effective as training data, we evaluated how the miss-detection rate on the real evaluation dataset changes with different ratios of real to synthetic images. The results are shown in Table 3. For this evaluation, we varied the ratio by increasing the number of real defective images in the training data using the same augmentation method described in the previous section. The results show that the miss-detection rate is strongly affected by the ratio of real to synthetic images. In this dataset, the lowest miss-detection rate was observed when the ratio was approximately 2:1 (real to synthetic). However, note that this optimal ratio may vary depending on the characteristics of the image dataset used.

Table 3. Change in miss-detection rate with different ratios of synthetic images in training

Training data (Images [Count])		Evaluation results (Miss-detection rate [%])
Real defect	Synthetic defect (Ratio of real to synthetic)	Real defect
500	1000 (1:2)	18.6
1000	1000 (1:1)	14.3
2000	**1000 (2:1)**	**7.2**
4000	1000 (4:1)	17.9

3-4 Effect of loop count

Figure 8 shows the performance of the defect detection model depends on the number of anti-weakness TLs. In this test, the ratio of real to synthetic images in the training data was fixed at 2:1, and the weakness extraction method described previously was employed. Zero TL count represents the initial condition, i.e., training performed using only real images without any synthetic images. The number of real defective images in the training data was 1,000. The miss-detection rate decreased significantly as the loop count increased: from 23.8% before the use of the synthetic image to 4.4% after the third loop. We also examined how the appearance property of the weakness

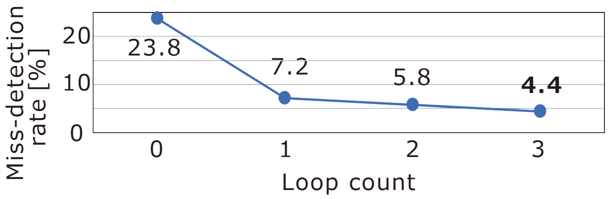


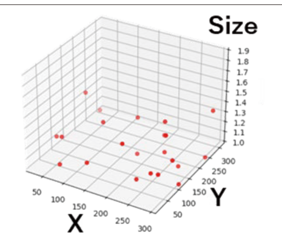
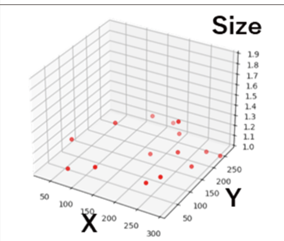
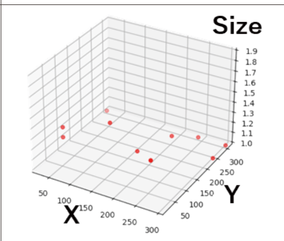
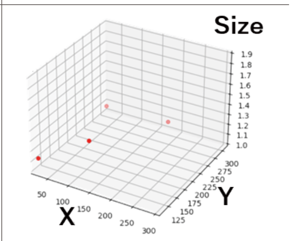
Fig. 8. Change in miss-detection rate with loop count

extracted in Step 4 changed with increasing loop count (Table 4). For visualization, the appearance properties identified as weaknesses were plotted as red dots in a three-dimensional space defined by defect location (x and y coordinates) and defect size. Initially, 22 weaknesses existed before starting the TL. However, this number gradually decreased with each loop. After the third loop, only four weaknesses remained. These results confirm that anti-weakness TL efficiently overcomes the weakness of the defect detection model and significantly improves its performance in terms of miss-detection rate.

3-5 Accelerating the development of defect detection model

For the ceramic material used in our experiment (shown in Fig. 6), the anti-weakness TL reduced the miss-detection rate to 4.4%, despite using only 10 original defective images (before augmentation) for training. In contrast, for another type of ceramic material, training a defect detection model using the conventional approach—without applying the anti-weakness TL—required 1,000 original real defective images to achieve a miss-detection rate below 5%. This demonstrates that the anti-weakness TL can reduce the necessary number of real defective images to 1/100. As noted in the introduction, acquiring real defective images on manufacturing site—where the defect rate is typically low—takes a considerable amount of time. Therefore, the ability to develop defect detection models using only a small number of real defective images significantly shortens the overall development period. In the case of this ceramic material used in our study, it was estimated that about four years would be necessary to acquire 1,000 real defective images. However, by applying the anti-weakness TL, including the training of image generation AI, a visual inspection system could be developed in about one month, resulting in a reduction of development time by over 90%.

Table 4. Change in the weakness of defect detection AI model with training loop count

Loop count	0th	1st	2nd	3rd
Weaknesses [Count]	22	15	10	4
Weakness distribution diagram				

4. Conclusion

We developed an anti-weakness TL using image generation AI to develop a defect detection model from a very small number of real images. This method enables the rapid development of a high-performance defect detection model using as few as approximately 10 real defective images, significantly reducing the time required for collecting training data. As a result, visual inspection systems can be developed in a much shorter period of time compared to conventional approach. In the future, we will extend the appearance property to support the extraction of more complex weaknesses, thereby broadening the range of products to which this technology can be applied.

Technical Terms

*1 Deep learning: One of the machine learning models. When training data is given, this model can automatically acquire the classification pattern. Although this model is known for its high classification performance, it has the shortcoming of requiring a large amount of training data.

*2 DATAGRID Inc.: A startup originating from Kyoto University, this pioneering company in generative AI has been engaged in its research and development since the early days of the technology, having been founded in 2017 when generative AI was still in its infancy.
HP: https://datagrid.co.jp/

*3 Image classification, object detection, and segmentation: Image classification refers to the task of assigning a class label to an entire image. Object detection involves identifying and locating objects within an image using bounding boxes. Segmentation is a more fine-grained task that detects object locations at the pixel level, providing a high-resolution delineation of object boundaries.

*4 Data leakage: Data leakage refers to a situation where information that should be unavailable during training is inadvertently included in the training data. When information related to evaluation or target variables leaks into the training process, the model may appear to perform well by effectively "knowing" the correct answers in advance. This leads to overly optimistic performance estimates and prevents proper evaluation of the model's true generalization ability.

References

(1) S. Shahinfar et al., ""How many images do I need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring," Ecological Informatics, 57, 101085 (2020)
(2) P. Bergmann et al., "The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection," International Journal of Computer Vision, vol. 129, no. 4, pp. 1038–1059 (2021)
(3) P. Bergmann et al., "MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9592-9600 (2019)
(4) A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929 (2020)
(5) K. He et al., "Deep Residual Learning for Image Recognition, CVPR2016, pp. 770–778 (2016)
(6) N. Kamali et al., "Characterizing Photorealism and Artifacts in Diffusion Model-Generated Images," arXiv preprint arXiv: 2502.11989 (2025)

Contributors

The lead author is indicated by an asterisk (*).

S. YOSHIDA*
• Assistant Manager, Digital Transformation Laboratory

S. ISHIDA
• Digital Transformation Laboratory

K. MIURA
• Ph.D. in Eng
Group Manager, Digital Transformation Laboratory