Semantic Reconstruction for Unified Detection of Local and Logical Anomalies

Kakeru Koizumi
Graduate School of FSE
Waseda University
Tokyo, Japan
kkeverio@ruri.waseda.jp

Taiga Hayami
Graduate School of FSE
Waseda University
Tokyo, Japan
hayatai17@fuji.waseda.jp

Hiroshi Watanabe
Graduate School of FSE
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—Anomaly detection is critical in industrial visual inspection, where undetected defects can lead to significant disruptions. While substantial progress has been made in detecting local anomalies such as scratches or dents, attention has recently shifted toward the more complex task of detecting logical anomalies, where visually normal components appear in semantically inconsistent configurations. To address this emerging task, we propose a novel model that resolves a key dilemma in Teacher–Student frameworks—namely, the risk of student overgeneralization to anomalous inputs. Our approach introduces a reconstruction-based student network that learns to restore object-level semantics, such as type and quantity, enabling a more holistic understanding of normality beyond localized cues. Experiments demonstrate that the proposed method achieves balanced performance on both local and logical anomalies.**

***Index Terms*—Anomaly Detection, image reconstruction, semantic segmentation.**

I. Introduction

Industrial visual inspection plays a crucial role in ensuring product reliability and safety. Traditional methods[1,2] have focused on local anomalies, including scratches, dents, and surface contamination. These methods, typically based on feature-space modeling or reconstruction-based learning, perform effectively under the assumption that anomalies appear as low-level pixel irregularities. However, with the increasing complexity and diversity of industrial environments, a new class of anomalies referred to as logical anomalies, as illustrated in Fig. 1, has emerged as an important research focus. Logical anomalies occur when visually normal components are arranged in semantically inconsistent ways. Detecting such inconsistencies requires an understanding of the object-level context rather than relying solely on texture or appearance.

In both local and logical anomaly detection, the scarcity of defective samples in real-world scenarios has made unsupervised learning[3,4] approaches predominant. Among these, Teacher–Student frameworks[5,6] have gained popularity because of their simplicity and effectiveness. In this setting, a student network is trained to replicate the intermediate features of a pretrained teacher network, and discrepancies between the two are interpreted as anomalies.

However, several studies [7, 8] have pointed out that training the student exclusively on normal data leads to overgeneralization, causing the network to produce representations that are overly similar to the teacher's, even for abnormal inputs—thereby degrading detection performance. To address this dilemma, the recent study [7] introduces denoising functionality based on autoencoder architectures into the student network. While such approaches have demonstrated effectiveness in handling local defects such as scratches or contamination, they remain limited in their ability to address logical anomalies arising from misplaced components or inconsistencies in object quantity.

Fig. 1. Example of a logical anomaly. Left: normal with two oranges and one apple. Right: anomaly with three apples, violating expected count.

To address the above limitations, we propose a novel anomaly detection framework based on semantic reconstruction at the object level. Unlike conventional approaches, the student network in our framework is not trained merely to replicate the teacher's features. Instead, it is supervised to reconstruct the correct configuration of objects, including their quantity, type, and spatial relationships. This is achieved by generating synthetic logical anomalies using segmentation and inpainting techniques. By training the model to restore plausible object-level semantics, we encourage a deeper understanding of structural normality. Our approach enables unified anomaly detection across both local and logical domains. Experimental results on the MVTec AD[9] and MVTec LOCO AD[10] datasets demonstrate that our method outperforms prior approaches in detecting local anomalies and achieves competitive performance on logical anomaly benchmarks as well.

979-8-3315-2415-9/25/$31.00 ©2025 IEEE

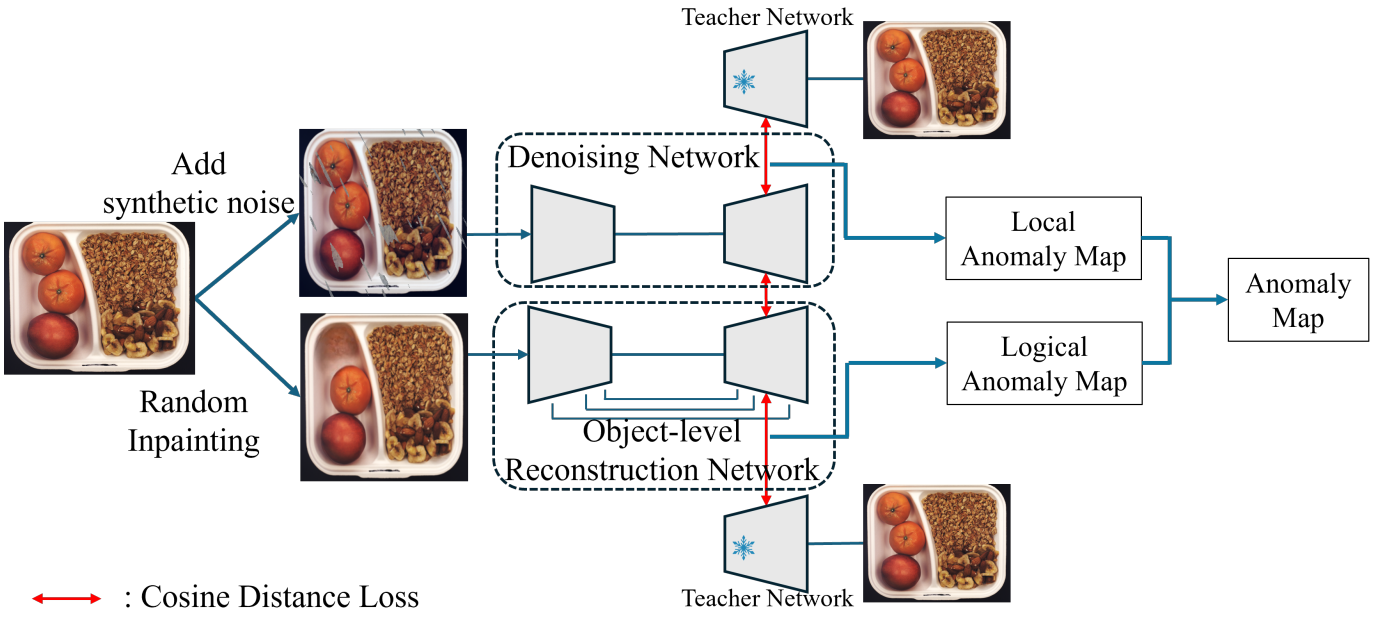

Fig. 2. Architecture of the proposed model, comprising a local anomaly removal module and an object-level reconstruction module. The student is trained on synthetic logical anomalies generated via SAM and inpainting.

II. Related works

A. Anomaly Detection Tasks

Traditional approaches to industrial anomaly detection have primarily focused on identifying local anomalies, such as surface scratches, dents, or contamination. These anomalies are typically characterized as pixel-level visual deviations from the normal distribution.Recently, methods that model the statistical distribution of features[11], employ self-supervised pretext tasks to learn normality representations[12], or perform nearest-neighbor search in a learned feature space[13] have been proposed. These methods detect anomalies based on either outlier scores, similarity metrics, or reconstruction errors from transformed inputs. However, recent research[14] has identified a new category of anomalies, referred to as logical anomalies, which involve components that appear visually normal but are arranged in semantically inconsistent ways. For example, an object may be present in the correct location but in an incorrect quantity. This type of anomaly is particularly relevant in industrial settings where spatial relationships and object counts are strictly defined. To address these challenges, To address these challenges, approaches that integrate both local texture cues and global contextual information have been proposed, with EfficientAD[8] serving as a representative example.

B. Unsupervised Teacher–Student Methods

Unsupervised Teacher–Student frameworks have become a dominant paradigm in visual anomaly detection due to their ability to operate without labeled anomalous data. In this setup, a student network is trained to replicate the intermediate feature representations of a fixed, pretrained teacher network using only normal samples. Anomalies are then inferred from discrepancies between the teacher and student outputs. However, one key limitation of this approach lies in the student's tendency to overfit or generalize excessively to the training distribution, which can result in reduced sensitivity to anomalies. To mitigate this issue, several enhancements have been proposed. Some approaches[15,16] restrict the learning capacity of the student network to prevent overfitting to the teacher's features. This is typically achieved by reducing the number of trainable parameters or introducing bottleneck layers in the student architecture, thereby encouraging sensitivity to out-of-distribution features. Others, such as DeSTSeg[7], incorporate an encoder–decoder structure that explicitly removes synthetic anomalies during training, allowing the student to focus on reconstructing clean representations and improving robustness to local anomalies.

III. Proposed Method

We propose a student model that reconstructs normal images from synthetically generated logical anomalies with object-level defects and local anomalies with additive noise. This design mitigates the optimization dilemma frequently observed in conventional Teacher–Student frameworks, where the student tends to overgeneralize to anomalous inputs. Unlike prior approaches, our method explicitly trains the student not merely to mimic the teacher's features, but to actively correct both local and logical anomalies. As a result, the student's reconstruction capability contributes directly to improved anomaly detection performance.

The overall architecture of the proposed model is illustrated in Fig. 2. It consists of two autoencoder-based subnetworks: a local anomaly removal module inspired by DeSTSeg, and an

TABLE I
QUANTITATIVE COMPARISON OF IAP (%) / IAP@90 (%) ON THE MVTEC LOCO AD DATASET (BEST IS IN BOLD, SECOND IS UNDERLINED)

Category	Method		
	DeSTSeg	**EfficientAD**	**Ours**
Breakfast Box	66.3 / **25.2**	63.0 / 11.3	**67.2** / 19.0
Juice Bottle	70.0 / 8.2	74.6 / 29.6	**74.8** / **36.0**
Pushpins	13.9 / 1.2	**16.1** / 1.0	14.0 / 2.0
Screw Bag	8.6 / 7.0	**15.6** / 4.0	12.0 / **7.4**
Splicing Connectors	24.7 / 8.7	**38.7** / **10.4**	22.7 / 7.8
Average	36.7 / 10.1	**41.6** / 11.3	38.1 / **14.4**

TABLE II
QUANTITATIVE COMPARISON OF IAP (%) / IAP@90 (%) ON THE MVTEC AD DATASET (BEST IS IN BOLD, SECOND IS UNDERLINED)

Category	Method		
	DeSTSeg	**EfficientAD**	**Ours**
Bottle	52.6 / 32.7	85.6 / 72.0	**91.5** / **86.4**
Cable	**92.5** / **83.9**	61.9 / 30.6	58.7 / 29.8
Capsule	**77.1** / **40.0**	57.9 / 36.7	55.3 / 36.0
Carpet	59.0 / 25.6	71.2 / 25.5	**76.1** / **31.9**
Grid	**82.4** /**56.8**	49.0 / 27.6	62.0 / 45.2
Hazelnut	86.4 / 77.1	65.4 / 53.0	**89.2** / **81.8**
Leather	**81.4** / 57.2	63.3 / 45.4	80.4 / **66.0**
Metal nut	41.8 / 20.6	84.6 / 80.2	**94.3** / **88.2**
Pill	**91.6** / **83.7**	73.4 / 65.1	78.5 / 34.9
Screw	**57.5** / **40.6**	44.2 / 15.0	50.3 / 10.3
Tile	87.7 / 78.5	76.6 / 60.8	**97.1** / **91.3**
Toothbrush	**73.3** / **60.6**	53.6 / 38.3	60.9 / 40.6
Transistor	**94.2** / **82.9**	89.3 / 81.3	81.8 / 61.2
Wood	86.6 / 69.4	63.7 / 44.3	**90.8** / **84.6**
Zipper	52.1 / 6.4	67.5 / 51.7	**85.3** / **70.1**
Average	74.4 / 54.3	67.1 / 48.4	**76.8** / **57.2**

object-level reconstruction module designed to restore semantic consistency. To enhance the student's ability to reconstruct object-level structures, we adopt a U-Net[17] architecture in the reconstruction branch.

To detect anomalies, we compute the cosine similarity between the decoder features of the two subnetworks and the encoder features of a pretrained ResNet-18 teacher model. Specifically, the local anomaly removal branch and the object-level reconstruction branch produce independent feature maps, which are respectively compared with the teacher's features to generate a local anomaly map and a logical anomaly map. These maps capture deviations at different semantic levels. The final anomaly map is obtained by summing the two, allowing for unified detection of both pixel-level and structure-level inconsistencies.

For training data generation, we begin with normal images and extract object regions using a combination of the Segment Anything Model (SAM) [18] and a foreground segmentation model, U²-Net[19]. Specifically, U²-Net generates a foreground mask, which is used as a spatial prior for SAM to extract semantically meaningful object candidates. One candidate is then randomly selected, and its corresponding region is modified using Inpainting Anything [20] to generate contextually coherent but semantically inconsistent object arrangements, simulating realistic logical anomalies.

IV. EXPERIMENTS

A. Set up

To validate the effectiveness and generalizability of the proposed method, we conduct experiments on two widely used benchmark datasets: MVTec AD, which contains 15 categories of industrial objects and textures primarily featuring local anomalies such as scratches, dents, and contaminations; and MVTec LOCO AD, which includes 5 categories specifically designed to evaluate logical anomalies, such as object miscounts, misplacements, and configuration inconsistencies.

Both the reconstruction and denoising branches are supervised using only normal images. The overall model is optimized to minimize the cosine similarity loss between the decoder features of the student network and the encoder features of the pretrained teacher model. We train with Adam optimizer, using a learning rate of 1e-4, batch size of 16, input resolution of 256×256, and up to 10000 steps with early stopping. All experiments are conducted on a single NVIDIA A4500 GPU.

As benchmark baselines, we use DeSTSeg [1] and EfficientAD [4]. Evaluation metrics are Image-level Average Precision (IAP) and IAP@90, where IAP measures general detection performance and IAP@90 focuses on high-confidence anomalies.

B. Results

The anomaly detection results on logical anomalies using the MVTec LOCO AD dataset are presented in Table 1. While our method slightly underperforms EfficientAD, which is specifically designed for logical anomaly detection, it surpasses DeSTSeg, one of the benchmark methods, in terms of average performance. Notably, in categories such as Breakfast Box and Juice Bottle, which are characterized by clear object-type and quantity features, our method achieves performance comparable to or even exceeding both baselines, demonstrating its effectiveness in capturing semantic consistency.

The results on local anomaly detection, evaluated on the MVTec AD dataset, are summarized in Table 2. Our method achieves the highest average IAP and IAP@90 across the dataset, outperforming both EfficientAD and DeSTSeg. In the category-wise comparison, it achieves the best IAP in 7 out of 15 categories, and the highest IAP@90 in 8 categories, showing strong and consistent performance across diverse types of local anomalies.

In contrast, for categories such as Capsule and Pill, which exhibit relatively simple structures, the proportion of the inpainted region tends to be large relative to the object

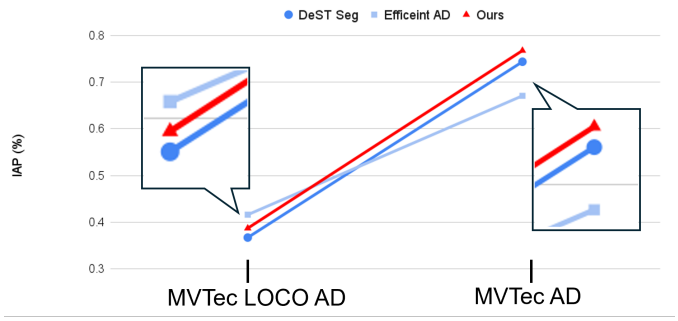

Fig. 3. Image-level IAP (%) on the MVTec AD and LOCO AD datasets.

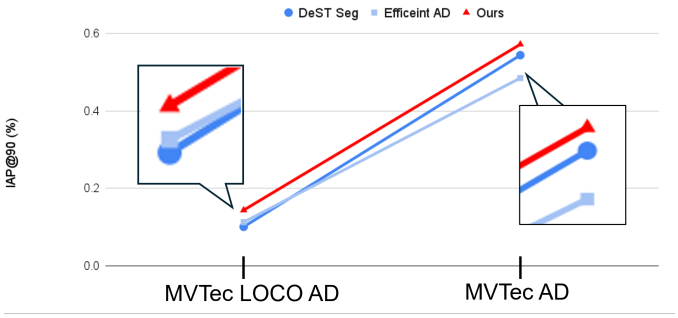

Fig. 4. Image-level IAP@90 (%) on the MVTec AD and LOCO AD datasets.

area extracted by SAM. This often leads to synthetic logical anomalies that deviate significantly from realistic distributions, potentially contributing to reduced detection performance.

Taken together, these quantitative results indicate that the proposed method offers well-balanced and competitive anomaly detection performance across both logical and local domains, often outperforming or matching state-of-the-art baselines depending on the anomaly type.

V. Conclusion

In this paper, we presented a novel anomaly detection framework that explicitly reconstructs object-level semantics to address both local and logical anomalies in industrial visual inspection. Unlike conventional Teacher–Student models, our method is designed to directly restore spatial and semantic consistency, thereby overcoming overgeneralization issues associated with anomalous inputs.

Through extensive experiments on the MVTec AD and MVTec LOCO AD datasets, our approach demonstrated superior or comparable performance across a wide range of categories. In particular, it showed remarkable effectiveness in scenarios that require understanding of structural relationships, such as object count or alignment. These results highlight the strength of incorporating semantic reconstruction into unsupervised anomaly detection. As future work, we plan to further improve detection robustness by synthesizing a wider variety of logical anomalies with higher fidelity.

References

[1] Z. Liu1, Y. Zhou, Y. Xu, and Z. Wang, "SimpleNet: A Simple Network for Image Anomaly Detection and Localization," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 20402-20411.

[2] M. Rudolph, B. Wandt, and B. Rosenhahn, "Same Same But DifferNet: Semi-Supervised Defect Detection with Normalizing Flows," IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021, pp. 1907-1916.

[3] D. Gudovskiy, S. Ishizaka, and K. Kozuka, "CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows," IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022, pp. 98-107.

[4] Y. Zheng, X. Wang, R. Deng, T. Bao, R. Zhao, and L. Wu, "Focus Your Distribution: Coarse-to-Fine Non-Contrastive Learning for Anomaly Detection and Localization," IEEE International Conference on Multimedia and Expo (ICME), 2022, pp. 1-6.

[5] T. Cao, J. Zhu, and G. Pang, "Anomaly Detection under Distribution Shift," IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 6511-6523.

[6] G. Wang, S. Han, E. Ding, and D. Huang, "Student-Teacher Feature Pyramid Matching for Anomaly Detection," Proceedings of the British Machine Vision Conference (BMVC), 2021.

[7] X. Zhang, S. Li, X. Li, P. Huang, J. Shan, and T. Chen, "DeSTSeg: Segmentation Guided Denoising Student-Teacher for Anomaly Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 3914-3923.

[8] K. Batzner, L. Heckler, and R. König, "EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies," IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 127-137.

[9] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD - A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9584-9592.

[10] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization," International Journal of Computer Vision, vol. 130, no. 4, pp. 947-969, 2022.

[11] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," European Conference on Computer Vision, 2020, pp. 475-492.

[12] C. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-Supervised Learning for Anomaly Detection and Localization," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 9659-9669.

[13] K. Roth, L. Pemula, J. Zepeda, B. Scholkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 14298-14308.

[14] J. Zhang, M. Suganuma, and T. Okatani, "Contextual Affinity Distillation for Image Anomaly Detection", IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 149-158.

[15] H. Deng, and X. Li, "Anomaly Detection via Reverse Distillation from One-Class Embedding," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 149-158.

[16] T. D. Tien1, A. T. Nguyen, N. H. Tran1, T. D. Huy1, S. T. M. Duong, C. D. Tr. Nguyen, S. Q. H. Truong, "Revisiting Reverse Distillation for Anomaly Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 24511-24520.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), 234–241, Oct. 2015.

[18] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 3992-4003.

[19] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane and M. Jagersand, "U²-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognition*, vol. 106, p. 107404, 2020.

[20] T. Yu, R. Feng, R. Feng, J. Liu, X. Jin, W. Zeng, and Z. Chen, "Inpaint Anything: Segment Anything Meets Image Inpainting," arXiv preprint arXiv:2305.18207, 2023.