합성된 이상 특징과 특화된 MLP 를 사용한 이상 탐지와 국소화

Anomaly Detection and Localization with Synthesized Anomaly Features and Assisted MLP

*,1 아미리안 바르누세파데라니 바하르, 1 바라시, 1 라크모노브아크로존, #,1 김정홍
1 경북대학교 컴퓨터학부
*,1Bahar Amirian Varnousefaderani, 1Barathi Subramanian, 1Rakhmonov Akhrorjon, and #,1Jeonghong Kim
1Department of Computer Science and Engineering, Kyungpook National University

ABSTRACT
Detecting and localizing anomalies in images are critical tasks across various domains. Applying machine learning and especially unsupervised methods has led to remarkable results in this task. However, the generalization of models to various domains and forms of anomalies is still improvable. This study presents an innovative approach that utilizes pre-trained models followed by a feature adaptor for feature preparation and synthesizes anomaly features by randomly adding three scales of Gaussian noise to the normal features. Then, two multilayer perceptrons (MLP), one used only for enhancing the other one in training time by affecting its loss, are utilized for the discrimination process. Experimental results show that compared to the state-of-the-art (SOTA) models, the proposed approach achieves superior anomaly detection with the highest AUROC for the largest number of objects (10 out of 15) on the MVTec AD dataset. It also outperforms on the Br35H dataset with an impressive AUROC of 99.5%, while providing remarkable results for localization on both datasets. Overall, our model exhibits enhanced generalization abilities while maintaining high efficiency compared to the existing baselines.

Key Words: Anomaly detection, Anomaly localization, Machine learning, Unsupervised learning

1. Introduction

Anomaly detection in images involves identifying anomalous samples, while anomaly localization refers to spotting specific areas containing anomalies. Anomaly detection plays a crucial role in various applications, including health assessment and ensuring the quality of industrial products. For instance, in health assessment, medical images are classified as either normal (healthy) or abnormal (indicating a possible illness), presenting a type of anomaly detection with the added challenge of localizing the problematic area. Such tasks become especially critical in situations where healthcare facilities and infrastructure are limited. In industrial settings, the ability to find and localize anomalies (defects) directly impacts quality and safety measures and drastically influences customer satisfaction. The conventional approaches for solving these tasks, however, are costly and error-prone due to the reliance on human labor. For this reason, automated visual inspections incorporating machine learning and deep learning techniques have been the subject of extensive research for years.

One important challenge in anomaly detection is that anomalies occurring on a single object can range from small scratches or slight changes in color to more severe defects or even component loss. This challenge becomes even more daunting when the goal is to develop an anomaly detection method that consistently leads to acceptable results for a variety of objects. Figure 1 illustrates an example of the normal form and various anomaly forms for two classes (tile and brain) from two distinct datasets (MVTec AD [1] and Br35H [2]) associated with two different domains (industrial and health).

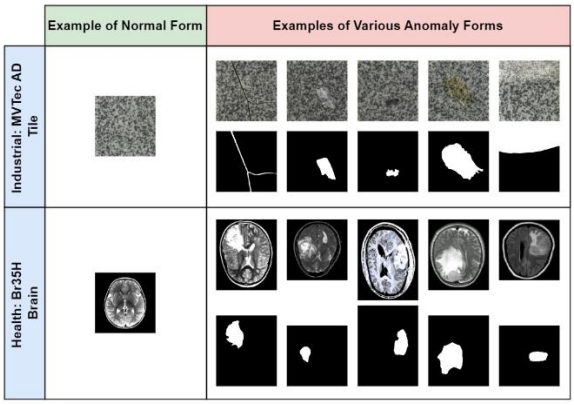

Fig. 1. Instances of normal form and various anomaly forms (upper line in each row) along with the corresponding ground-truths (lower line in each row) for the 'tile' and 'brain' classes from the MVTec AD [1] and Br35H [2] datasets related to the industrial and health domains respectively. Our model outperforms baseline models in anomaly detection for both classes, achieving an AUROC of 99.9% for tile and 99.5% for brain.

This intrinsic diversity, coupled with the scarcity of anomalies, make constructing large and comprehensive datasets including a sufficient number of all types of anomalies exceedingly difficult, thereby hindering the effectiveness of supervised methods. This has led to a growing interest in unsupervised techniques that exclusively rely on normal samples to address the problem.

Unsupervised approaches for anomaly detection and localization can generally be divided into two main categories: reconstruction-based [3], [4] and out-of-distribution-based [5]–[10] methods. Reconstruction-based methods are developed

*발표자(bahar.am.va@gmail.com), #교신저자(jhk@knu.ac.kr)

based on the argument that a model trained solely on normal samples lacks the ability to accurately reconstruct anomalies. Consequently, the difference between the original sample and its reconstruction serves as the indicator of anomalous behavior. On the other hand, out-of-the-distribution-based methods involve identifying samples that deviate from normal samples or their distribution.

Although state-of-the-art (SOTA) results have been achieved using such methods, especially with the utilization of pre-trained models, there is still room for improving the generalization of models across various input domains and anomaly forms.

In this paper, the process of training an efficient cross-domain model for detecting and localizing anomalies begins by extracting normal features. This is achieved by passing the normal images through a pre-trained model followed by a feature adaptor that alleviates the domain bias. Next, anomalous features are generated by randomly adding three scales of Gaussian noise to the normal features, representing various forms and types of anomalies. Finally, with both normal and anomalous features available, a multilayer perceptron (MLP) is innovatively trained to discriminate anomalies while dealing with the randomness introduced in the last step, leading to improved generalization. This is accomplished by employing a simpler assistive MLP which influences the loss of the main, more complex MLP during training. Testing and inference are conducted by simply extracting the features of input images and passing them through the discriminator.

To summarize, this work makes the following contributions:

1. Generation of anomaly features by randomly adding three scales of Gaussian noise to the normal features.
2. Introduction of an innovative approach for training MLP as the discriminator.
3. Demonstration of improved generalization capabilities of the proposed technique through experiments conducted on two datasets from distinct domains.

This relatively simple yet efficient model adeptly handles the task, showcasing enhanced robustness and superior performance across the majority of classes due to the mentioned design choices.

2. Related Work

Recent unsupervised anomaly detection and localization methods can be broadly categorized into two primary groups: Reconstruction-based and Out-of-Distribution-based methods. These groups are further explained in this section.

2.1 Reconstruction-Based Methods

Reconstruction-based methods are designed around the idea that a reconstruction model trained only with normal samples will struggle to accurately reconstruct anomalous regions. Detection and localization are achieved by measuring the pixel-wise error between the original and reconstructed samples.

The reconstruction process can be achieved through various techniques. A popular approach involves the utilization of generative models with the latest techniques being founded on diffusion models [11] as seen in [4] which is inspired by the ability of latent diffusion models [12] to generate high-quality and diverse images.

A similar strategy involves reconstruction-by-inpainting as illustrated in RIAD [3] where partial image regions are randomly removed and the purpose is reconstructing the image from partial inpaintings.

Despite the logical appeal of these methods, many of them unintentionally learn to reconstruct anomalies due to over generalization which leads to failure in the systems.

2.2 Out-of-Distribution-Based Methods

Out-of-distribution-based methods are employed to identify samples that deviate from the distribution of the normal data. In the broader context presented in this paper, these methods assess the degree of abnormality by comparing the features of test samples with a set of normal patch features or by comparing the test samples with normal samples using the normal distribution learned during training.

Extracting features in such methods is accomplished through a range of techniques found in various studies. Some works rely on self-supervised methods for feature extraction as illustrated in CutPaste [5] where the pretext task involves cutting an image patch and pasting it at the random location of a larger image.

Several other methods achieve SOTA results by employing deep networks pre-trained on large datasets such as ImageNet [13] to extract generalized normal features, as seen in PaDim [6], RevDist [7], SimpleNet [8], and patchCore [9]. Among these methods, PaDim [6] leverages multivariate Gaussian distributions to create a probabilistic representation of normal features but has limitations in performance. RevDist [7], instead, utilizes the concept of training a smaller and more compact model to match the behavior of a larger one, a technique commonly known as knowledge distillation also found in [10]. The dissimilarity between the two models is considered as anomaly in this approach, but the challenge persists with nearly doubled computational requirements.

Despite the strong results offered by these methods, a lack of adaption leads to a mismatch between the high-level features extracted from ImageNet [13] and those required in other domains. PatchCore [9] addresses this issue by utilizing mid-level features to minimize the bias towards ImageNet [13] classes and aggregates over a local neighborhood to keep sufficient spatial context. SimpleNet [8] further enhances the results by implementing a simple feature adaptor and framing the problem as a binary classification task. However, there is room for improvement, particularly when facing a variety of types, sizes, and forms of anomalies in various domains.

3. Proposed Methodology

The methodology employed in this paper comprises three primary components as illustrated in Figure 2: Feature Preparation, Anomalous Feature Generation (only for training), and Discrimination. These components are described in details in this section.

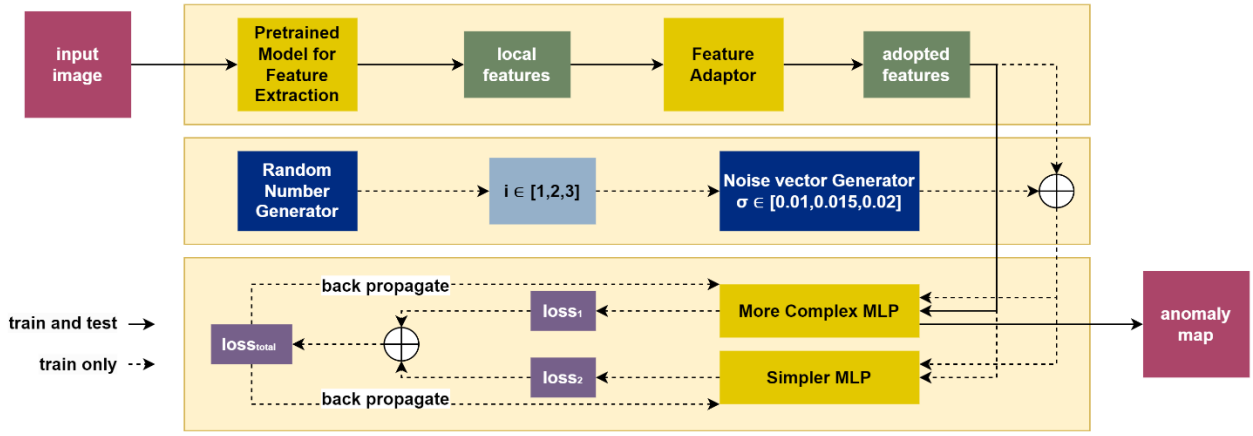

Fig. 2. Overview of the proposed method. Each of the three main components of the method are illustrated in a separate row. During training, normal features are achieved by passing the normal samples through a pre-trained model for feature extraction followed by a feature adaptor for decreasing the impact of domain bias. Anomalous features are achieved by the random addition of three scales of Gaussian noise to the normal features. The normal and anomalous features are then used for training a discriminator comprising a more complex main MLP assisted by a simpler MLP. For testing and inference, the anomalous feature generation component is removed and samples are passed through the feature preparation and discrimination components.

3.1 Feature Preparation

Preparing the features is achieved in two steps. The first step involves extracting local features in a similar approach to [9] and the second step refers to adapting the acquired features to the target domain similar to [8]. The second step helps in improving the results as ImageNet [13] usually has a different distribution with the images found in anomaly detection datasets.

More specifically, each image x_i in train and test set is passed through a ResNet-like pertained network denoted as ϕ and a subset $\mathcal{S}$ of its extracted feature map levels are used for the next stages. Each of these feature maps are denoted as $\phi_s(x_i) \in \mathbb{R}^{H_s \times W_s \times C_s}$ where $s \in \mathcal{S}$ and H_s, W_s, and C_s are the height, width, and channel size of the feature map respectively. In each of the feature maps, a feature slices at position $h \in H_s$ and $w \in W_s$ is denoted as $\phi_s(x_i, h, w) \in \mathbb{R}^{C_s}$ and its incorporation with the neighborhood features with patch size p is denoted as

$$\mathcal{N}_p^{(h,w)} = \left\{(n,m) \mid n \in \left[h - \left\lfloor\frac{p}{2}\right\rfloor, \dots, h + \left\lfloor\frac{p}{2}\right\rfloor\right], \; m \in \left[w - \left\lfloor\frac{p}{2}\right\rfloor, \dots, w + \left\lfloor\frac{p}{2}\right\rfloor\right]\right\} \quad (1)$$

where n and m are the height and width of this neighborhood. Then, adoptive average pooling is used for the aggregation of features in the neighborhood $\mathcal{N}_p^{(h,w)}$. Denoting this aggregation function as f_{agg}, the local feature $\mathcal{o}_s(x_i, h, w)$ is defined as

$$\mathcal{o}_s(x_i, h, w) = f_{agg}\left(\{\phi_s(x_i, n, m) \mid (n,m) \in \mathcal{N}_p^{(h,w)}\}\right). \quad (2)$$

All feature maps from the chosen levels are then resized to the size of the largest one and are concatenated to give $g(x_i)$ where each of its entries at location h, w is defined as $g(x_i, h, w)$.

In the next step, a feature adaptor $\mathcal{A}$ is implemented as a single fully-connected layer without bias to project the features to the target domain and gives

$$\mathcal{k}(x_i, h, w) = \mathcal{A}(g(x_i, h, w)). \quad (3)$$

3.2 Anomalous Feature Generation

To train a discriminator capable of distinguishing between positive (normal) and negative (anomalous) samples, it is common to collect both normal and anomalous samples. However, due to the scarcity of anomalous samples in anomaly detection scenarios, it becomes necessary to generate such samples. Although anomalous samples are typically generated by synthesizing anomalies directly onto the normal images, generating anomalous features can lead to improved results [8].

In this paper, anomalous features are generated by randomly adding three scales of Gaussian noise to the normal features. The scale of the added noise determines the strength of the anomalous features [8]. Larger noise scales result in a significant difference between normal and anomalous features, making it challenging for the discriminator to identify less severe anomalies. Conversely, very small noise scales may hinder the discriminator from converging effectively. Therefore, employing three scales of noise enables the discriminator to identify various forms of anomalies without setting overly loose or tight boundaries.

During each step in the training process, a random number $i \in [1,2,3]$ is generated for each normal image x_i in the train set. Based on this number, one of the three predefined noise scales $\sigma \in \{0.01, 0.015, 0.02\}$ is chosen. Then, a vector $\epsilon \in \mathbb{R}^C$ with entries sampled from the i.i.d. Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ is added to the features of each normal image to give

$$\mathcal{k}_-(x_i, h, w) = \mathcal{k}(x_i, h, w) + \epsilon \quad (4)$$

which is considered as a sample of anomalous features.

3.3 Discrimination

This component is explained separately for training and for testing and inference.

For Training, both positive and generated negative

samples are fed into two MLPs simultaneously: a more complex main MLP denoted as $\mathcal{D}_1$ and a simpler assistive MLP denoted as $\mathcal{D}_2$. Both of these discriminators assign a normality score to each location h, w producing normality as $\mathcal{D}_i(\mathcal{h}(h,w))$ where $i \in \{1,2\}$. The discriminators expect positive values for normal features and negative values for anomalous features. In accordance with this expectation, the loss is computed as follows. Each of the discriminators consider two thresholds, th and $-th$, and apply a $l1$ loss, linearly penalizing normal features scored below th and anomalous features scored above $-th$. Subsequently, the total loss l_{total} is calculated as the sum of the losses computed by each model. l_{total} is then used to train both models through backpropagation.

For testing and inference, first the feature preparation process is applied on each image. The extracted features are then fed to $\mathcal{D}_1$ only and the produced $-\mathcal{D}_1(\mathcal{h}(h,w))$ serves as the anomaly scores needed for anomaly localization (using anomaly map) and anomaly detection (using anomaly detection score). Anomaly map is composed by interpolating and applying a Gaussian filter on the anomaly scores and anomaly detection score is the maximum among these scores.

This design choice is based on the argument that a shallow network performs well when synthesized anomalies are closely related and the problem is less complex. However, a deeper model is needed when the problem is more complex. During training, as the complexity of the problem at each stage is randomly defined as a result of the randomness introduced by the anomalous feature generation stage, the cooperation of these two models enhances the results by applying more updates based on their points of agreement and having less impact on the model for points of disagreement.

4. Experimental Results

The details of datasets, evaluation metrics, implementation details, experiments, discussions, and results are presented in this section.

4.1 Datasets

The experiments are conducted on two datasets: MVTec AD [1] and Br35H [2].

MVTec AD is an industrial dataset including 5354 images in total categorized into 5 textures (namely carpet, grid, leather, tile, and wood) and 10 objects (namely bottle, cable, capsule, hazelnut, metal nut, pill, screw, toothbrush, transistor, and zipper) classes. This dataset consists of normal samples for training and both normal and various defective samples for testing. The pixel-level annotations of defects are also provided which is used as the ground-truth for anomaly localization.

Br35H originally consists of 1500 tumorous and 1500 non-tumorous brain MRI images. Segmentation maps are also provided for 800 of the tumorous images. To make sure that the dataset is balanced, 1250 non-tumorous images are randomly selected for training. Also, 250 non-tumorous images as normal samples and 250 tumorous images with their segmentation maps as anomalous samples are randomly selected to be used for testing.

4.2 Evaluation Metrics and Baselines

Area under the receiver operating characteristic (AUROC) is employed to measure the performance of the proposed method for both detection and localization.

More specifically, image-level AUROC is calculated using the produced anomaly detection scores for evaluating anomaly detection. On the other hand, pixel-level AUROC is gained using the anomaly map for evaluating anomaly localization.

For comparing the performance of the proposed method on MVTec AD [1], RIAD [3], CutPaste [5], PaDim [6], RevDist [7], and SimpleNet [8] are chosen as the baselines. For comparison on Br35H [2], the two best competitors on MVTec AD [1], namely RevDist [7] and SimpleNet [8] are selected as the baselines.

4.3 Implementation Details

The implementation setup for the experiments of this paper is as follows.

The feature extractor in the first step of feature preparation is a WideResnet50 which is pre-trained with ImageNet [13]. Similar to [9] and [8], the 2nd and 3rd intermediate layers extracted using this network are chosen as the effective levels to decrease the bias towards ImageNet [13] while maintaining a high performance and efficiency. Also, the feature dimension from feature extractor is set as 1536. In the second step of feature preparation, the feature adaptor is a fully connected layer without bias similar to the selected structure in [8].

In the discrimination component, the simpler assistive MLP consists of one set of a linear layer, batch normalization, and a leaky ReLU with a slope of 0.2, while the more complex main MLP comprises three of such sets, and both are followed by a liner layer. The threshold th used for calculating the losses is 0.5 and $-th$ is -0.5.

Adam optimizer with a learning rate of 0.0001 for the adaptor and 0.0002 for the discriminator is used for training in addition to a weight decay of 0.00001. The training procedure consists of 160 epochs for each class with a batch size of 4.

The implementations were carried out using Python 3.8 and PyTorch library version 1.12.1 utilizing one 8 GB NVIDIA GeForce RTX 2060 SUPER GPU with CUDA 11.

4.4 Results on MVTec AD

Anomaly detection and localization results on MVTec AD [1] are shown in Table 1 for comparison with the baseline models using the format of Image-wise AUROC/Pixel-wise AUROC for each entry (The format remains consistent for the next experiments).

The proposed method achieves the highest image-wise AUROC score for the largest number of classes (10 out of 15) compared to the baseline models. Impressively, 7 of these classes achieve a perfect image-wise AUROC score of 100%. This method also results in competitively high pixel-wise AUROC scores for all classes and, on average, demonstrating its effectiveness in anomaly localization.

4.5 Results on Br35H

The experiment is repeated on Br35H [2] for comparing the proposed method with the two models that achieved the best competitive results on MVTec AD [1].

Table 1. Comparison of our model with the SOTA works on MVTec AD [1]. Results are shown in the format of Image-wise AUROC/Pixel-wise AUROC for each entry.

Type	Reconstruction-based	Out-of-distribution-based				
Model / Class	RIAD [3]	CutPaste [5]	PaDim [6]	RevDist [7]	SimpleNet [8]	Ours
Carpet	84.2/96.3	93.9/98.3	**99.8**/**99.1**	98.9/98.9	99.7/98.2	99.6/98.4
Grid	99.6/98.8	**100**/97.5	96.7/97.3	**100**/**99.3**	99.7/98.8	**100**/98.6
Leather	**100**/99.4	**100**/**99.5**	**100**/99.2	**100**/99.4	**100**/99.2	**100**/99.2
Tile	98.7/89.1	94.6/90.5	98.1/94.1	99.3/95.6	99.8/**97.0**	**99.9**/96.1
Wood	93.0/85.8	99.1/**95.5**	99.2/94.9	99.2/95.3	**100**/94.5	99.7/95.0
Bottle	99.9/98.4	98.2/97.6	99.1/98.3	**100**/**98.7**	**100**/98.0	**100**/98.0
Cable	81.9/84.2	81.2/90.0	97.1/96.7	95.0/97.4	**99.9**/**97.6**	99.0/97.2
Capsule	88.4/92.8	**98.2**/97.4	87.5/98.5	96.3/98.7	97.7/98.9	97.3/**99.1**
Hazelnut	83.3/96.1	98.3/97.3	99.4/98.2	99.9/**98.9**	**100**/97.9	**100**/98.2
Metal Nut	88.5/92.5	99.9/93.1	96.2/97.2	**100**/97.3	**100**/**98.8**	**100**/98.5
Pill	83.8/95.7	94.9/95.7	90.1/95.7	96.6/98.2	99.0/**98.6**	**99.1**/98.5
Screw	84.5/98.8	88.7/96.7	97.5/98.5	97.0/**99.6**	**98.2**/99.3	96.9/99.3
Toothbrush	**100**/98.9	99.4/98.1	**100**/98.8	99.5/**99.1**	99.7/98.5	**100**/98.4
Transistor	90.9/87.7	96.1/93.0	94.4/97.5	96.7/92.5	**100**/**97.6**	**100**/96.6
Zipper	98.1/97.8	99.9/99.3	98.6/98.5	98.5/98.2	**99.9**/**98.9**	**99.9**/98.8
Average	91.7/94.2	96.1/96.0	95.8/97.5	98.5/97.8	99.6/98.1	99.4/98.0

The main motivation for this experiment is evaluating the performance of the methods in distinct domains and while facing a wider range of anomaly forms, types, and sizes. It is worth mentioning that this problem is even more challenging, as the variations among normal images of brain in this dataset are significantly higher than the normal samples available in the MVTec AD [1] dataset.

The results of this experiment are shown in Table 2. The proposed method achieves the best result for image-wise AUROC when discriminating tumorous and non-tumorous images of brain. The observation that SimpleNet [8], the best competitor when comparing the results on MVTec AD [1], falls behind in this experiment, demonstrates the superior generalization of our model across various domains and forms of anomalies.

Table 2. Comparison of our model with the two best competitors on MVTec AD [1] for Br35H [2]. Results are shown in the format of Image-wise AUROC/Pixel-wise AUROC for each entry.

Model / Class	RevDist [7]	SimpleNet [8]	Ours
Brain Tumor	99.2/**92.4**	98.6/86.5	**99.5**/92.0

4.6 Discussion

Further experiments are conducted to demonstrate the effectiveness and necessity of the innovative technique employed for training the discriminator.

Without employing this technique, one could utilize a traditionally trained single MLP. It is observed that in this case, optimal results are obtained by applying a 4-layer MLP as it yields the best results in most of the cases and displays fewer signs of under-fitting and over-fitting in comparison with other structures. This observation leads to choosing a 4-layer MLP as the more complex main MLP in our technique. However, the randomness introduced by the anomalous feature generation process not only constrains the performance of such methods in comparison to the SOTA models, but also complicates the task of identifying the most suitable structure for all cases. This challenge motivates the implementation of a technique to assist the single MLP in various situations, as proposed in this paper.

Given this understanding, to specifically illustrate the effectiveness of applying a simpler MLP that influences the loss of the main MLP during training and is discarded at test time, the model is compared with two other versions: 1. Keeping the main structure of the discrimination component but utilizing the simpler MLP as the main discriminator during test and inference (Ours_V2), and 2. Utilizing only a traditionally trained single 4-layer MLP as the discriminator (Ours_V3).

As shown in table 3, the proposed method outperforms the other version for most of the classes and exhibits better overall performance in both anomaly detection and localization due to the employment of the explained technique.

Table 3. Comparison of our selected model comprising a more complex MLP assisted with a simpler MLP (Ours_Selected) with two other versions: 1. Using the simpler MLP as the main MLP (Ours_V2) and 2. Traditionally training the more complex MLP without assistance from the simple MLP (Ours_V3). Results are shown in the format of Image-wise AUROC/Pixel-wise AUROC for each entry.

Model / Class	Ours_Selected	Ours_V2	Ours_V3
Carpet	**99.6**/98.4	99.1/97.1	99.4/**98.5**
Grid	**100**/98.6	97.7/98.6	99.7/**98.7**
Leather	**100**/**99.2**	**100**/**99.2**	**100**/**99.2**
Tile	99.9/**96.1**	99.6/94.4	**100**/94.6
Wood	99.7/**95.0**	98.4/74.1	**100**/89.3
Bottle	**100**/98.0	**100**/96.9	**100**/**98.1**
Cable	**99.0**/**97.2**	96.4/96.5	**99.0**/96.9
Capsule	**97.3**/**99.1**	96.8/98.6	97.1/99.0
Hazelnut	**100**/**98.2**	**100**/97.6	**100**/**98.2**
Metal Nut	**100**/**98.5**	99.8/98.2	**100**/**98.5**
Pill	**99.1**/**98.5**	97.5/97.8	98.3/**98.5**
Screw	**96.9**/**99.3**	86.7/98.5	96.2/99.1
Toothbrush	**100**/98.4	97.5/98.3	**100**/**98.6**
Transistor	**100**/**96.6**	**100**/90.6	**100**/96.1
Zipper	**99.9**/**98.8**	99.7/98.5	99.6/**98.8**
Brain Tumor	**99.5**/92.0	97.4/88.9	98.9/**92.2**
Average	**99.4**/**97.6**	97.9/95.2	99.3/97.1

4.7 Qualitative Results

Exemplary samples illustrating anomaly localization on various classes of the MVTec AD [1] and Br35H [2] datasets are visualized in Figure 3 and Figure 4 respectively. The threshold for segmentation results is established by evaluating the F1-score across all anomaly scores associated with each sub-class following [8].

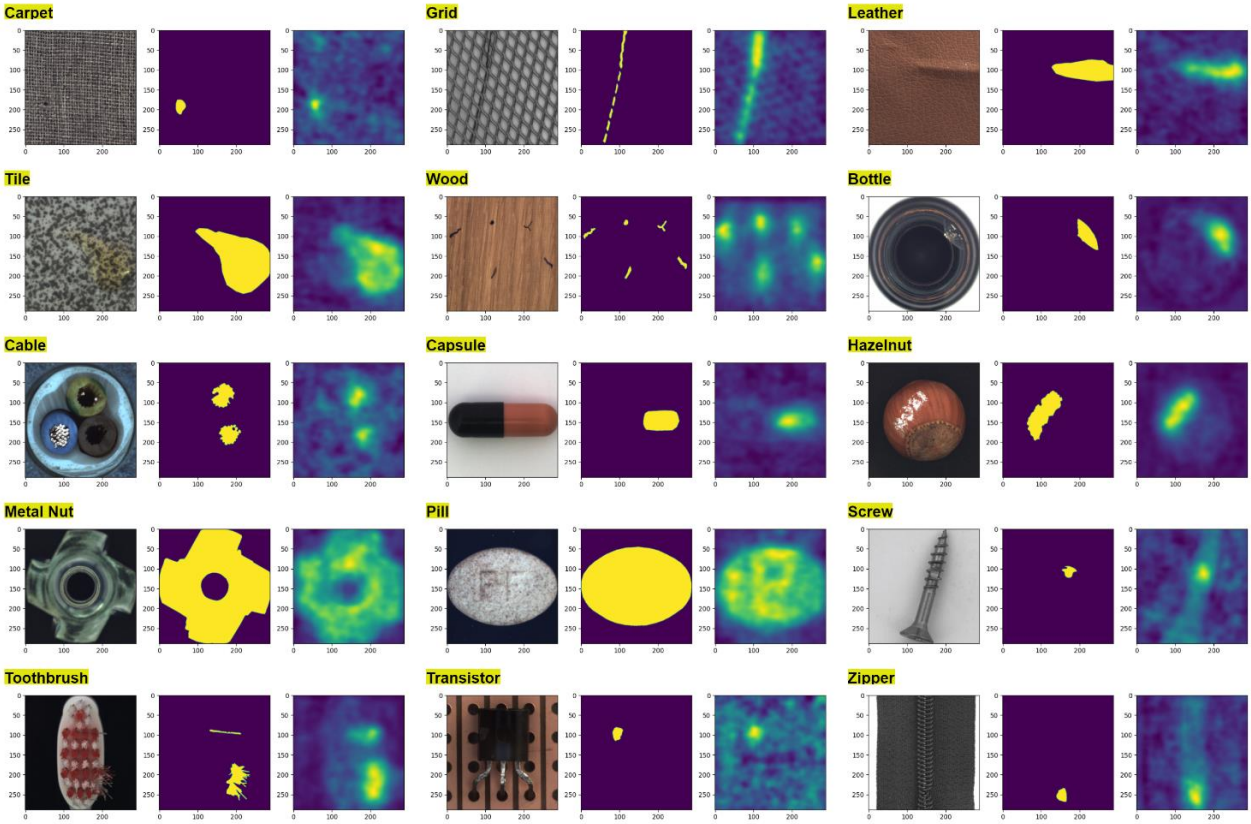

Fig. 3. Illustration of qualitative results for each of the classes in the MVTec AD [1] dataset. Each entry showcases an exemplary anomaly sample of one of the classes in this dataset, its corresponding ground truths, and its generated anomaly map from left to right. The ability of the proposed method to effectively localize anomalies with diverse forms, types, and sizes is evident even within this limited set of samples.

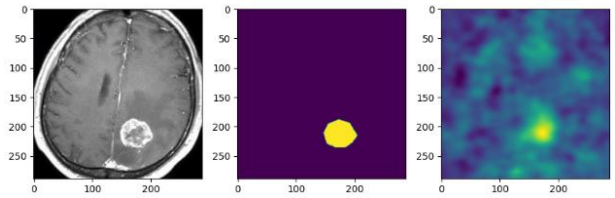

Fig. 4. Illustration of qualitative results for the Br35H [2] dataset, featuring an exemplary anomaly sample, its corresponding ground truth, and its generated anomaly map from left to right.

5. Conclusion

A relatively simple yet effective solution to the problem of image anomaly detection and localization across diverse domains was introduced in this paper. This method involves utilizing a pre-trained model followed by a simple feature adaptor for feature preparation, generating representative anomaly features by randomly adding three scales of Gaussian noise to normal features, and innovatively implementing a discriminator with assisting MLPs. The experimental results affirmed the effectiveness of the proposed method in detecting and localizing a diverse range of anomaly forms, types, and sizes across all classes in two datasets selected from different domains, exhibiting competitively high performance compared to SOTA models.

Limitations and Future Work. While the generalization of this method is evaluated on two datasets, a widely used industrial dataset and a health assessment dataset, additional experiments could be conducted on larger-scale datasets to further assess its performance across a broader array of domains and classes.

Moreover, while the proposed method has been characterized as relatively simple yet effective, it is essential to recognize the necessity of continuous efforts in refining its capabilities. There exists potential for enhancements in various aspects of this method, including efficiency, faster inference times, and reduced resource utilization.

Acknowledgments

This study was supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394).

References

1. Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger, 2019, MVTec AD - A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection, P*roceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9592-9600.

2. A. HAMADA, 2022, Br35H :: Brain Tumor Detection, https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
3. Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj, 2021, Reconstruction by inpainting for visual anomaly detection, *Pattern Recognition*, 112, pp. 107706.
4. Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, Shu-Tao Xia, 2023, Unsupervised Surface Anomaly Detection with Diffusion Probabilistic Model, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6782-6791.
5. Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister, 2021, CutPaste: Self-Supervised Learning for Anomaly Detection and Localization, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9664-9674.
6. Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier, 2021, PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization, *International Conference on Pattern Recognition*, pp. 475-489.
7. Hanqiu Deng and Xingyu Li, 2022, Anomaly Detection via Reverse Distillation from One-Class Embedding, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9737-9746.
8. Zhikang Liu, Yiming Zhou, Yuansheng Xu, Zilei Wang, 2023, SimpleNet: A Simple Network for Image Anomaly Detection and Localization, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402-20411.
9. Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Scholkopf, Thomas Brox, and Peter Gehler, 2022, Towards Total Recall in Industrial Anomaly Detection, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318-14328.
10. Akhrorjon Akhmadjon Ugli Rakhmonov, Barathi Subramanian, Bekhzod Olimov, and Jeonghong Kim, 2023, Extensive Knowledge Distillation Model: An End-to-End Effective Anomaly Detection Model for Real-Time Industrial Applications, *IEEE Access*, 11, pp. 69750-69761.
11. Jonathan Ho, Ajay Jain, and Pieter Abbeel, 2020, Denoising Diffusion Probabilistic Models, *Advances in neural information processing systems*, 33, pp. 6840-6851.
12. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer, 2022, High-Resolution Image Synthesis with Latent Diffusion Models, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684-10695.
13. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, 2009, ImageNet: A Large-Scale Hierarchical Image Database, *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255.