2024년도 한국인터넷정보학회 추계학술발표대회 논문집 제25권 2호

조건 기반 디노이징 확산 모델을 활용한 텍스처 이상 탐지

아미리안 바르누세파데라니 바하르*, 라크모노브아크로존*, 김태훈**, 김정홍*
*경북대학교 컴퓨터학부, **딥비젼
{bahar.amirian.v, r.akhror}@knu.ac.kr, dipvision.ceo@gmail.com, jhk@knu.ac.kr

Conditioned-Guided Denoising Diffusion Model for Texture Anomaly Detection

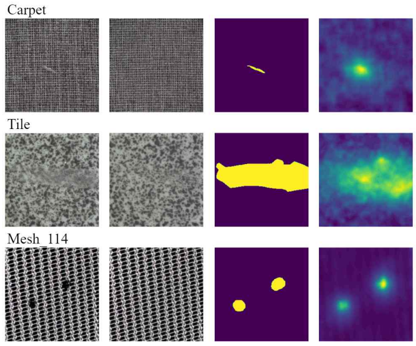
Bahar Amirian Varnousefaderani, Rakhmonov Akhrorjon, Taehun Kim, Jeonghong Kim

1. Introduction

Anomaly detection (AD) is a crucial task across various domains, including industrial manufacturing for ensuring the quality of products. In this domain, AD involves identifying anomalous samples, and anomaly localization refers to pinpointing the specific anomalous areas within those samples. These tasks are particularly challenging due to the scarcity and diversity of anomalies. As a result, creating comprehensive datasets that encompass all possible types of anomalies is not feasible, and supervised techniques are often impractical. Consequently, most recent methods are unsupervised, relying only on normal data for training.

Unsupervised methods can be broadly categorized into representation-based methods [1] and reconstruction-based methods [2]. Representation-based methods define normality and treat anything that deviates from this as anomaly. Reconstruction-based methods, alternatively, train a generative model on normal samples, learning their distribution. Since the model struggles with reconstructing anomalies, they are detected by comparing the input with its anomaly-free reconstruction. However, previous methods in this area face two significant issues: unintentionally generating anomalies or producing low-quality reconstructions leading to misidentification.

In this paper, we focus on texture AD and localization and utilize diffusion models [3] as the generative model to develop an effective reconstruction-based approach. Unlike conventional usage of these models, our approach starts the sampling process with a partially diffused sample rather than pure noise, using conditioned sampling and blur guidance to improve reconstruction quality. We also combine feature-wise and pixel-wise comparisons between the original and reconstructed samples for more accurate results. Our method surpasses state-of-the-art techniques in AD and localization on the texture classes of the MVTec benchmark [4] and representative classes from the DTD-Synthetic dataset [5]. Figure 1 visualizes the anomaly localization results for exemplary samples.



(Figure 1) (Anomaly localization visualization for exemplary samples. For each class, the original sample, the reconstructed version, the ground truth, and the created anomaly map are shown in order.)

2. Related research

2.1 Method

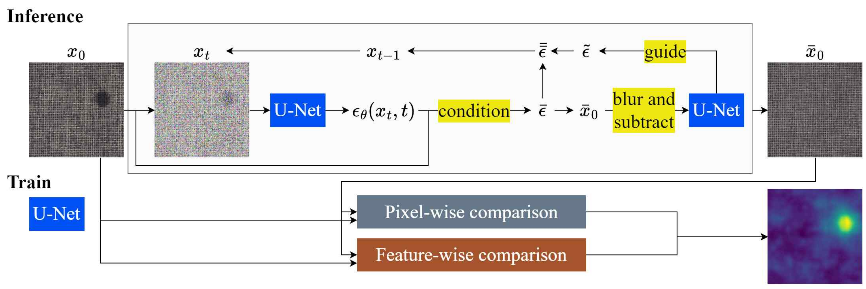
We propose a reconstruction-based method for unsupervised texture AD using diffusion models. Figure 2 illustrates an overview of our method

Denoising diffusion models are generative models that create images from noise through an iterative denoising process. For our problem, a diffusion model is first trained on normal data, similar to conventional diffusion models.

During inference, the goal is to reconstruct each test sample without anomalies using three main strategies. First, the reconstruction process begins with a partially noised version of the input. This approach encourages the model to focus on the key features of the input sample while ignoring any potential anomalous details.

Second, the reconstruction is conditioned. In this step, the score function is conditioned on the original sample, and the initial error is updates to reflect the conditioning.

Finally, the reconstruction quality is improved using the blur guidance technique from [6], which leverages Gaussian blur to guide the diffusion process without external inputs. Here, we compute the difference between the original prediction and its blurred version. This difference is then passed through the trained model to generate an additional error, which is subsequently used to update the original error for further refinement.



(Figure 2) (An overview of the method. The training phase involves training a U-Net to predict the noise at each time stamp using only normal data. The inference phase consists of reconstructing an anomaly-free version of the input and effectively comparing the original and reconstructed versions.)

With the original sample and its reconstructed version, the next step is to perform an effective comparison between the two and generate an anomaly map. This comparison is carried out in two ways: a basic pixel-wise comparison and an additional feature-wise comparison performed by comparing the features of the original sample and its reconstructed version.

2.2 Experiments

We conduct our experiments on two main datasets: the texture classes of the MVTec dataset [4] and representative classes from the DTD-Synthetic dataset [5], to compare the AD and localization performance of our method with state-of-the-art methods. The results are shown in Tables 1. Additionally, the anomaly localization results for exemplary samples are visualized in Figure 1.

(Table 1) (Comparison of the performance of our method with that of the representative state-of-the-art methods. Sample-wise AUROC / Pixel-wise AUROC are presented for each element in percentage.)

Type	Representation -based	Reconstruction -based	
Method	SimpleNet [1]	DRAEM [2]	Ours
MVTec [4]			
Carpet	99.7/98.2	97.0/95.5	99.2/98.9
Grid	97.2/98.8	99.9/99.7	100/99.2
Leather	100/99.2	100/98.6	100/99.3
Tile	99.8/97.0	99.6/99.2	100/98.2
Wood	100/94.5	99.1/96.4	100/98.2
DTD-Synthetic [5]			
Mesh_114	99.2/98.5	97.1/98.0	99.8/98.5
Stratified_154	100/98.6	100/99.3	100/99.3
Blotchy_099	100/99.0	100/99.2	100/99.4

3. Conclusion

We introduced a promising reconstruction-based method for texture AD and localization using a conditioned-guided denoising diffusion model. To address challenges such as inadvertently reconstructing anomalies while maintaining high-quality outputs, we incorporated conditioned reconstruction and blur guidance, along with starting the sampling process from a partially diffused input. Additionally, we utilized a combination of pixel-wise and feature-wise comparisons between the original and reconstructed samples to effectively localize anomalies.

Future work includes expanding experiments to more classes, larger datasets, and additional objects. We plan to integrate and calibrate the attention-based guidance from [6] for AD to enhance the results. We will also detail the mathematical formulations and present results from our ablation studies.

Acknowledgement

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 (4120240214871), and also by the project for Collabo R&D between Industry, University, and Research Institute funded by the Korea Ministry of SMEs and Startups in 2023 (RS-2023-00221775).

References

[1] Liu, Z., Zhou, Y., Xu, Y., Wang, Z., “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, 2023.

[2] Zavrtanik, V., Kristan, M., Skocaj, D., “Draem-a discriminatively trained reconstruction embedding for surface anomaly detection”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8330 - 8339, 2021.

[3] Ho, J., Jain, A., Abbeel, P., “Denoising diffusion probabilistic models”, Advances in neural information processing systems, 33, 6840 - 6851, 2020.

[4] Bergmann, P., Fauser, M., Sattlegger, D., Steger, C., “Mvtec ad – a comprehensive real-world dataset for unsupervised anomaly detection”, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9592 - 9600, 2019.

[5] Aota, T., Tong, L.T.T., Okatani, T., “Zero-shot versus many-shot: Unsupervised texture anomaly detection”, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5564 - 5572, 2023.

[6] Hong, S., Lee, G., Jang, W., Kim, S., “Improving sample quality of diffusion models using self-attention guidance”, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7462 - 7471, 2023.