2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Application of Denoising Image Restoration to Anomaly Detection

Yu Kashihara† and Takashi Matsubara†

†Graduate School of Engineering Science, Osaka University
1–3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
Email: kashihara@hopf.sys.es.osaka-u.ac.jp, matsubara@sys.es.osaka-u.ac.jp

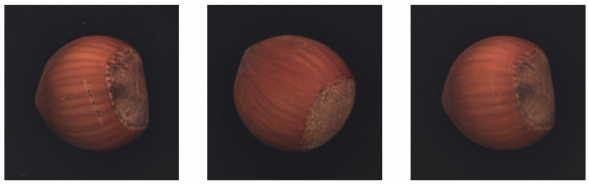
Abstract—Generative models learn complicated distributions and generate new samples that follow the learned distributions. Approximating the input image with the learned model is called reconstruction. Anomaly detection by generative models is achieved by comparing the reconstruction and original images. However, existing generative models often lose the original features. The anomaly detection often fails if the reconstruction loses the original features. We propose the anomaly detection model based on the diffusion model to avoid this problem by simple method. In this study, the model is evaluated on MVTeC AD, a dataset of industrial products anomaly detection, and demonstrates the area under receiver operating characteristic curve of 0.92. The score is significantly better than the existing generative models. We also show that the denoising model restores anomalous regions by the proposed method. The most significant contribution of our method is that they outperform existing reconstruction methods by using the diffusion model trained in the usual way and using simple reconstruction method.

1. Introduction

Anomaly detection aims to recognize outliers of the data point or unexpected patterns from a dataset or its feature space. It has been widely used in different applications, especially in computer vision, such as tumor identification in medical images [1], industrial defects detection [2], and road traffic monitoring [3]. With booming in deep learning, numerous advanced frameworks in anomaly detection have been proposed, and the results show the promise [4, 5]. However, the practically labeling procedure of anomaly samples is a time-consuming task, driving the anomaly detection methods away from the supervised learning manner. The anomaly detection community needs a more general framework that can escape the teaching signal limitation.

Deep generative models (DGM) recently became the new fashion in the unsupervised learning community, which aims to learn the explicit data distributions using deep neural networks and then generate new data from this distribution. Various DGM-based anomaly detection frameworks are proposed to compare the input image with the reconstruction, where the reconstruction can be viewed as an approximation for the input image.

ORCID iDs First Author: 0000-0002-7191-100X, Second Author: 0000-0003-0642-4800



(a) (b) (c)

Figure 1: Effects of partial transitions on the reconstruction with DDPM. (a)raw image (b)using whole transitions (c)using partial transitions

To name a few, variational autoencoder (VAE) [6] and generative adversarial network (GAN) [7] are popularly designed as the backbone of anomaly detection and have exhibited high performance. However, DGM-based anomaly detection requires the model to have a powerful reconstructed ability for each input image to prevent the loss of original key features. That is always a crux for VAE due to its strong assumption (normal Gaussian) of latent data distribution and too powerful autoregressive decoder [8]. Meanwhile, GAN-based models inevitably lose original features, such as orientation and detail flows [9].

Denoising diffusion probabilistic model (DDPM) [10] is a recently proposed DGM inspired by Langevin dynamics that can generate high-quality reconstructions [10]. DDPM consists of two phases: *diffusion process* and *reverse process*. In the diffusion process, the original input image is gradually transformed into the noise by adding noises, while in the latter process, the model removes the noises and reconstructs the image with the reverse sequential step to diffusion. In practice, the anomaly image can be viewed as one type of noise image comparing the normal image. We can implement anomaly detection if the model removes the noise region and identifies where it is denoised. Motivated by this, this paper proposes a DDPM transformation that implements the anomaly detection by DDPM with partial transitions (example as illustrated in Figure 1). In a normal DDPM transition, the diffusion process completely transforms the image into noise, and the original features are lost. On the other hand, DDPM with partial transitions preserves the original features because the diffusion process does not completely transform the image into noise.

For proof-of-concept, a challenging task, i.e., industrial anomaly detection, is demonstrated by introducing MVTeC AD dataset [11]. The experimental results show that the proposed method can preserve the original orien-

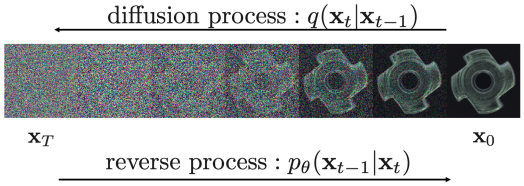


Figure 2: Schematic of DDPM. $q(\mathbf{x}_t|\mathbf{x}_{t-1})$ is a true distribution of the forward process. $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is a model predicting true distribution of the reverse process $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$.

tation, shape, and performance superior to existing generative models. Furthermore, the proposal successfully recovers anomalous regions in anomalous images.

2. Denoising Diffusion Probabilistic Models

As shown in Figure 2, DDPM is a Markov chain that first transits each input image to a destroyed noise (*diffusion* in Section 2.1), then produces a noise sample and reconstructs it to match the input after a finite time (*reverse* in Section 2.2) [10].

2.1. Diffusion process

The diffusion process is the transitions in which raw sample $\mathbf{x}_0$ follows $q(\mathbf{x}_0)$ turns into noisy samples $\mathbf{x}_1, ..., \mathbf{x}_T$, and these transitions follow the Gaussian distribution with the variance schedule β_t. This process is regarded as a Markov chain evolving according to Gaussian noise. The transition from sample $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ is represented by the following:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t;\ \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \tag{1}$$

$$q(\mathbf{x}_{1:T}) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}). \tag{2}$$

Sample $\mathbf{x}_t$ can be represented in closed form by the reproductive property of the Gaussian, and the equation is represented as follows using the notations $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_s^t \alpha_s$.

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t;\ \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t))\mathbf{I}. \tag{3}$$

Here, the diffusion process does not have learnable parameters because the transition only depends on the variance schedule β_t.

2.2. Reverse process

The reverse process follows the distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$, gradually removing the noise ϵ. However, it is difficult to calculate the true posterior. For this reason, we use an approximation to predict the true distribution following bellow.

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \beta_t\mathbf{I}). \tag{4}$$

The means and variances of the model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ are learnable parameters. The model uses the variance schedule β_t same as the diffusion process, while the mean $\mu_\theta(\mathbf{x}_t, t)$ is predicted by denoising models ϵ_θ.

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)\right). \tag{5}$$

The model at the final step above is regarded as a decoder, which outputs $\hat{x}_0$ as:

$$\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\alpha_1}}\left(\mathbf{x}_1 - \sqrt{\beta_1}\epsilon_\theta(\mathbf{x}_1, 1)\right). \tag{6}$$

The denoising model ϵ_θ is trained by minimizing the following l_1 norm for any step t:

$$L(\theta) = \mathbb{E}_{t,\mathbf{x}_t,\epsilon}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|_1\right]. \tag{7}$$

3. Proposed method

Proposed method utilizes the partial transitions of the full transitions used for training to reconstruct images. In the diffusion process with the full transitions over T steps, the input image becomes almost noisy at T steps.

We propose a method composed of two tricks. The first trick is to perform the partial diffusion process on the input image, followed by the reverse process with the same number of steps. The trick performs a diffusion process from the input image $\mathbf{x_0}$ to the over t steps according to Equation 2 and reconstructs it from $\mathbf{x_t}$ using Equation 4. The partial diffusion process allows reconstruction while preserving original image features.

The second trick regards the input image as an image containing noise and reconstructs it only by the reverse process without diffusion. The trick considering an image including anomaly as a sample $\mathbf{x}_t$ to which noise has already been injected, our proposal ideally utilizes the reverse process to denoising or fix the anomaly region without adding additional noise by the diffusion process. We treat the input image $\mathbf{x}_0$ as $\mathbf{x}_t$, assuming an anomalous image that contains noise. The denoising model generates images by removing noise starting from the input image, not the diffusioned noise image, while the denoising model ϵ_θ is expected to capture the anomalies as noise in $\mathbf{x}_t$ and reconstruct them.

These two tricks can reconstruct the anomalous areas of the input image while preserving the normal areas. This paper uses the squared error of $\mathbf{x}_t$ and $\hat{x}_0$ as a pixel-wise anomaly score. The normal regions almost do not change much, so the anomaly score is small. On the other hand, the anomalous regions change a lot over transitions, so the anomaly score is large.

To determine whether an image includes an anomaly or not, we propose an image-wise anomaly score to maintain that the measurements of all images are normalized. Input image size is (Width, Height, N_{ch}) where N_{ch} is the number

of channels. The image-wise anomaly score is the maximum value of the average pooling of an image over the channel direction by the kernel. Here, the kernel size is (W_k, H_k, N_k), and the stride is S. Average pooling can convert pixel-wise anomaly scores into patch-wise ones while allowing the proposed method to capture a wider range of image features.

4. Experiments and Results

4.1. Preparation

As a proof-of-concept, we examine the efficacy of the proposed method on the industrial anomaly problem. Specifically, we compare the proposed platform against related algorithms on authoritative datasets MVTeC AD [11]. MVTeC AD includes 15 classes: ten object classes and five texture classes whose minimum size of the image is 700×700 pixels, and the maximum is $1,024 \times 1,024$ pixels, respectively. We preprocess for alignment of the images, and the final images were resized to 192×192 pixels used in our experiments. The training data sets contain only normal images, while the test sets contain both normal and anomalous images.

The models were trained using randomly rotated and flipped images. This processing is one of the data augmentations that can reduce overfitting on the model [12].

4.2. Implementation details

In experiments, we set timestep to $T = 1,000$ to be the same as the model of the original DDPM [10]. For the variance of the diffusion process, we used the cosine variance schedule proposed by Nichol et al. [13]. DDPM with the cosine variance schedule achieved better sampling quality than ones with linearly increasing variance schedule [13]. The cosine variance schedule is represented as:

$$\bar{\alpha}_t = \frac{f(t)}{f(0)},\ f(t) = \cos\left(\frac{t + Ts}{T + Ts} \cdot \frac{\pi}{2}\right). \tag{8}$$

Here, s is a small offset to prevent β_t from becoming an infinitesimal value.

The model uses the U-Net-based PixelCNN model [14] with group normalization and self-attention throughout to represent the reverse process. The DDPM models for each class were trained in 300000 steps, each with a batch size of 32. The number of steps of the transitions in the proposed method is empirically set to 5 steps for the texture category and ten for the object category. The kernel size of the average pooling was set to a size that rounded off 1/25 of the input image size. Finally, we evaluated anomaly detection performance by AUROC at the image level.

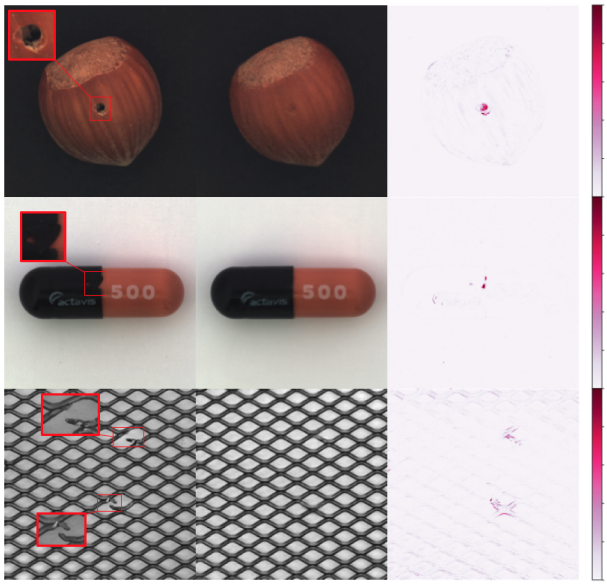
4.3. Results

We compared the proposed method with other DGM models and the DDPM with full transitions. Here, GANomaly [5] and ARNet [4] were used as comparison methods. GANomaly's results were taken from [9], and ARNet's results were taken from [15].

Table 1: AUROC on MVTeC AD

	Class	ARNet	GANomaly	DDPM	Proposed method	
					diffusion	no diffusion
Texture	Grid	0.88	0.71	0.68	**1.00**	**1.00**
	Leather	0.86	0.84	0.92	**0.99**	**0.99**
	Tile	0.74	0.79	0.65	**0.94**	**0.94**
	Carpet	0.71	0.70	0.42	0.46	**0.74**
	Wood	0.92	0.83	0.83	**0.98**	0.97
	Average	0.82	0.77	0.71	0.91	**0.93**
Object	Bottle	0.94	0.89	0.78	**0.99**	**0.99**
	Capsule	0.68	0.73	0.42	**0.90**	**0.90**
	Pill	0.79	0.74	0.67	**0.84**	**0.84**
	Transistor	0.84	0.79	0.57	0.93	**0.94**
	Zipper	0.88	0.75	0.63	0.93	**0.94**
	Cable	**0.83**	0.76	0.58	**0.83**	**0.83**
	Hazelnut	0.86	0.79	0.79	**1.00**	**1.00**
	Metal Nut	0.67	0.70	0.40	0.87	**0.88**
	Screw	**1.00**	0.75	0.43	0.81	0.81
	Toothbrush	**1.00**	0.94	0.69	**1.00**	**1.00**
	Average	0.85	0.76	0.60	**0.91**	**0.91**
All average		0.84	0.64	0.70	0.91	**0.92**

GANomaly is an anomaly detection model combined with GAN and autoencoder. GANomaly consists of three sub-networks: autoencoder, encoder, and discriminator. GANomaly is learned to minimize a composite loss function of three networks. The reconstruction is obtained as an output of the autoencoder network. The anomaly score is related to latent variables of the reconstruction. ARNet consists of attribute erasing module (AEM) and the attribute restoration network (ARNetwork). AEM erases the semantic features of the original image, and ARNetwork aims to restore the partially erased image to minimize the difference from the original image. In addition, the results were also compared with DDPM using full transitions. The method using the full transitions is indicated as DDPM, and the proposed method using the diffusion process is indicated as diffusion, and without diffusion is indicated as no diffusion. Table 1 shows anomaly detection performance by these methods. In most classes, the proposed method demonstrated the best anomaly detection performance. These results show that the proposed method demonstrates excellent performance in reconstruction-based methods. The performance of the proposed method without diffusion is a little better than that of the diffusion method in most classes, indicating the usefulness of using only the reverse process for reconstruction. Compared to existing generative models [5, 4], the AUROC of the proposed method for the Capsule, Metal Nut, and Hazelnut classes have significantly been improved to 0.90, 0.88, and 1.00, respectively. These classes possess image-specific features such as rotation and character print. This result suggests that the proposed method succeed in preserving the original features. The proposed method also improved AUROC in the texture category. The results show that the proposed method is effective for images with repetitive patterns. On the other hand, the anomaly detection performance did not improve much for the carpet and screw classes. These classes contain minimal anomalies. One



(a) original (b) reconstruction (c) squared error

Figure 3: Restoration of anomalous images without diffusion. The number of reverse steps is set to the optimal number of steps for each. (a) Original images. (b) Reconstruction images. (c) The squared error of the original image and the reconstruction image. The values are summed in the channel direction and divided by their maximum value for normalization [0,1] to visualize the anomaly score.

possible reason is that the input of the proposed method was resized to 192×192 pixels, so the anomalies' resolution was insufficient. Several other experiments were conducted to investigate the cause of the performance degradation caused by the proposed method, but a clear cause of the performance degradation caused was not found.

Next, we show the restoration results of anomalous images in Figure 3 by the method without the diffusion process. The reconstruction restores the anomalous regions included in the original image. The anomaly score is calculated by taking the original and reconstructed images' squared error and adding them together in the channel direction. The visualization of the squared error also shows that anomalous regions are extracted. The results show that original features such as fiber direction are preserved in the hazelnut in the first row and character position in the capsule in the second row. The denoising model captures the anomalous regions as noise and restores them, preserving original features by setting the reverse steps to optimal.

5. Conclusion

We tackled anomaly detection using DDPM and demonstrated excellent performance among generative models by simple reconstruction. The new anomaly detection method that DDPM uses partial transitions can reconstruct images while preserving original image features. As a result, the model achieved an AUROC of 0.92, which outperformed comparison models. The anomalies were also successfully visualized using anomalous regions restoration.

Acknowledgment

This work was supported by JST, PRESTO (JPMJPR21C7) and JSPS KAKENHI (19H04172, 19K20344, 21H03515), JST-Mirai Program (JPMJMI20B8) Japan.

References

[1] B. H. Menze, A. Jakab *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Trans Med Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[2] T. Defard, A. Setkov *et al.*, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," *LNCS*, p. 475–489, 2021.

[3] Y. Li, J. Wu *et al.*, "Multi-Granularity Tracking with Modularlized Components for Unsupervised Vehicles Anomaly Detection," *CVPR Workshops*, pp. 2501–2510, 2020.

[4] C. Huang, F. Ye *et al.*, "Attribute Restoration Framework for Anomaly Detection," 2019.

[5] S. Akcay, A. Atapour-Abarghouei *et al.*, "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training," *LNCS*, p. 622–637, 2019.

[6] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *ICLR*, 2014.

[7] I. Goodfellow, J. Pouget-Abadie *et al.*, "Generative Adversarial Nets," in *NeurIPS*, 2014, pp. 2672–2680.

[8] G. Perarnau, J. van de Weijer *et al.*, "Invertible Conditional GANs for Image Editing," in *NeurIPS Workshop*, vol. abs/1611.06355, 2016.

[9] F. Carrara, G. Amato *et al.*, "Combining GANs and Autoencoders for Efficient Anomaly Detection," *ICPR*, 2021.

[10] J. Ho, A. Jain *et al.*, "Denoising Diffusion Probabilistic Models," in *NeurIPS*, 2020, p. 6840–6851.

[11] P. Bergmann, M. Fauser *et al.*, "MVTec AD — A Comprehensive Real-world Dataset for Unsupervised Anomaly Detection," in *CVPR*, 2019, pp. 9584–9592.

[12] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *ArXiv*, vol. abs/1712.04621, 2017.

[13] A. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," in *ICML*, vol. 139, 2021, pp. 8162–8171.

[14] T. Salimans, A. Karpathy *et al.*, "Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications," 2016, pp. 1747–1756.

[15] M. Rudolph, T. Wehrbein *et al.*, "Fully Convolutional Cross-Scale-Flows for Image-based Defect Detection," 2021.