Exploring different noise distributions in diffusion models for Surface Anomaly Detection

Matic Fučka, Vitjan Zavrtanik, Danijel Skočaj

University of Ljubljana, Faculty of Computer and Information Science
Email: maticfuc@gmail.com, {vitjan.zavrtanik, danijel.skocaj}@fri.uni-lj.si

Abstract

Diffusion models have demonstrated considerable promise in the realm of image generation, receiving substantial research attention. Due to their powerful high-fidelity reconstruction capabilities, they have been shown to have potential as a foundation of reconstruction-based surface anomaly detection methods. While numerous studies have investigated the significance of diverse noise distributions within diffusion models, their exploration in the context of Surface Anomaly Detection has remained limited. To address this, we conduct an investigation of the impact of several distinct noise distributions on the anomaly detection capability of diffusion-based reconstructive anomaly detection models. The results of our study provide critical insights into the choice of noise distributions, encouraging further exploration of distributions that emulate real-world defects.

1 Introduction

Surface anomaly detection is a critical task in the field of industrial inspection, as it aims to identify abnormal images and accurately localize anomalous regions [1, 2]. The application of conventional supervised learning approaches, however, is challenging due to the scarcity of labeled abnormal images. Existing surface anomaly detection methods bypass this issue by leveraging only anomaly-free data to localize anomalies effectively.

Surface anomaly detection can currently be efficiently addressed through various methods. Discriminative methods focus on learning the boundary that separates normal and abnormal samples through synthetic anomaly generation during training [3, 4, 5]. Embedding-based methods [6, 7] use features extracted by pretrained backbones to build a memory bank [8] or fit a Gaussian distribution [9, 10, 11] over the data. Reconstruction-based methods [12, 13, 14, 15] accurately reconstruct only anomaly-free regions of the image while restoring the abnormal regions closer to the normal appearance.

Diffusion models have been shown to be effective image generation models [16]. Diffusion models gradually refine an image by progressively removing noise. Diffusion-based surface anomaly detection methods [17] rely on the standard Gaussian noise distribution and have not significantly explored the design space of diffusion processes. A few have explored other noise distributions
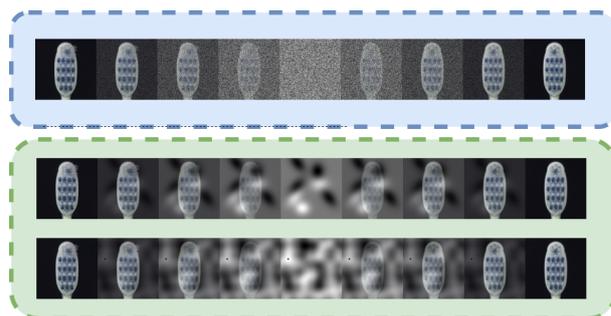

Figure 1: While Diffusion models typically use Gaussian noise it may benefit the reconstruction process if a noise distribution with an unequal frequency specter was used. The default diffusion process is shown in blue, a noise distribution with an unequal frequency specter process is visualized in green. Each row showcases the diffusion process (transition from the first toothbrush image to noise) and the reverse process (transition from noise to the second toothbrush image). The three visualized noises are samples from the three noise distributions that are explored in this study.

[18] but were not yet applied to the industrial inspection domain. A recent study [19] has demonstrated that alternative noise distributions can be employed effectively, indicating that the Gaussian distribution is not a necessary requirement for diffusion models.

Due to this, we hypothesize that alternative noise distributions may be beneficial for surface anomaly detection and that noise distributions that have an unequal frequency specter may enable the model to better learn the normal appearance of an object. We evaluate the impact of various noise distributions on the surface anomaly detection performance of a diffusion-based anomaly detection model. Our best model that utilizes Gaussian noise distribution achieves 93.0% I-AUROC on MVTec AD benchmark [1] and beats most of the previously proposed reconstruction methods.

2 Related work

Surface anomaly detection Surface anomaly detection has emerged as a highly active research area in recent years, with numerous approaches proposed to tackle this challenging problem. Initially, early methods employed auto-encoders (AE) [20, 12, 13, 21] or generative adver-

sarial networks (GANs) [14, 15] to reconstruct objects and detect surface anomalies. However, a shift in research focus has been observed, where the use of raw images has transitioned to the utilization of feature maps [7, 8, 22, 9] extracted from pre-trained networks, enabling the learning of normality based on these feature representations. Additionally, an alternative research path has emerged, involving the creation of synthetic anomalies [3, 4, 5] and learning another discriminative network to discriminate between the normal and abnormal images.

Diffusion models Diffusion models have experienced a resurgence in the past two years, following the initial work of Ho *et al.* [16], which introduced the basic diffusion model. The main idea behind their diffusion model is the reverse process, which slowly denoises the initially noisy image to a clear image. The approach has been successfully extended to various domains, such as audio and text generation [23, 24, 25, 26].

Given that the reverse process of diffusion models can be computationally expensive, recent work has focused on improving the efficiency and effectiveness of these models. Various improvements have been proposed, such as changes to the loss function, noise distributions, and sampling procedures [27, 28, 29, 30, 31]. An interesting approach was proposed by Bansal *et. al.* [19], who showed that diffusion models can invert arbitrary image transformations without the necessary use of Gaussian noise, but can be done with completely deterministic image degradations such as blur, masking, and more.

Unsupervised anomaly detection with diffusion models Wyatt *et. al.* [18] proposed AnnoDDPM which is based on the original diffusion model but instead of adding Gaussian noise they added Simplex noise. They applied this method to a dataset of medical images and achieved state-of-the-art results, but did not apply the process to the industrial inspection domain.

Zhang *et. al.* [17] proposed a DRAEM[3]-like network (composed of the reconstructive and segmentation network) where the reconstruction network is a standard DDPM and the segmentation network is an UNet. They make two steps at various points in time and average both of the segmentation masks. The reconstructive network is a DDPM which they did not alter in any way and they are making single-step reconstructions.

3 Diffusion models for unsupervised anomaly detection

3.1 Diffusion models

Diffusion models [23, 26, 27, 16], a subclass of neural networks, belong to the family of generative models and have been inspired by non-equilibrium thermodynamics [32]. These models are composed of three main components: a diffusion process, a reverse process, and a score model. The diffusion process, also known as the forward process, generates a sequence of noisy observations from a target distribution, while the reverse process generates

samples from a initial distribution. The score model is responsible for estimating the steps in the reverse process.

In the DDPM the forward process is defined by progressively adding Gaussian noise for T time-steps. More precisely we can parameterize each step as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \quad (1)$$

where $\beta_1, ..., \beta_T$ is the variance schedule and are held constant as hyperparameters in the DDPM. Due to this formulation, it can be easily seen that we can sample x_t at any timestep t with the following formula:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I), \quad (2)$$

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \epsilon\sqrt{1-\bar{\alpha}_t}, \ \epsilon \sim \mathcal{N}(0,1), \quad (3)$$

where $\alpha_i = 1 - \beta_i$ and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$.

During training a UNet-like [33] model $\epsilon_\theta(x_t, t)$ is trained to predict ϵ by minimizing the ℓ_2 loss:

$$\mathcal{L} = \mathbb{E}_{t\sim[1-T],x_0\sim q(x_0),\epsilon\sim(N)(0,I)}\Big[||\epsilon - \epsilon_\theta(x_t,t)||^2\Big]. \quad (4)$$

At inference stage x_{t-1} is reconstructed from of x_t with the following formula:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\Big(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\Big) + \sigma_t z, \quad (5)$$

where $\sigma_t^2 = \beta_t$ and $z \sim \mathcal{N}(0,1)$. x_0 is reconstructed from x_T iteratively from $x_T \to x_{T-1} \to ... \to x_0$.

3.2 Anomaly Detection

Anomaly mask M is derived from two sources: the original image I and the reconstructed diffused image x_0. To calculate x_0 we first diffuse I for T_0 steps to receive the diffused image x_{T_0}. The resulting diffused image x_{T_0} is then iteratively reconstructed towards x_0 with the usage of the reverse process. Once we have x_0 the anomaly mask M is calculated as:

$$M = ||I - x_0||_2^2. \quad (6)$$

The anomaly mask M is then postprocessed with a mean filter of size $n \times n$. The final anomaly score is calculated as the maximum value of the anomaly mask M.

4 Proposed Noise Distributions

We evaluated several different noise distributions to determine their impact on the surface anomaly detection performance of diffusion models. In Sections 4.1, 4.2 and 4.3, the Gaussian noise distribution, the Simplex noise distribution and the mixture of Gaussian Mixture Model distribution are described, respectively.

4.1 Gaussian noise

In the case of Gaussian noise, each pixel in the image is sampled independently from the standard normal distribution, denoted as $\mathcal{N}(0,1)$.

By sampling noise from this distribution, which is added to each pixel, a random and uncorrelated perturbation is introduced, allowing for a straightforward and accessible representation of noise in the context of this study.

Method	Recon. Methods				State-of-the-art			Ours		
	[34]	[14]	[35]	[36]	[3]	[8]	[9]	Norm.	Simpl.	Gauss Mix. Model
Carpet	84.2	69.9	70.6	98.8	97.0	98.7	**100**	76.4	93.5	87.4
Grid	99.6	70.8	88.3	**100**	99.9	98.2	99.7	**100**	93.8	94.0
Leather	**100**	84.2	86.2	**100**	**100**	**100**	**100**	99.6	99.5	99.9
Tile	98.7	79.4	73.5	98.2	99.6	98.7	**100**	98.5	99.4	99.3
Wood	93.0	83.4	92.3	97.5	99.1	99.2	**100**	98.2	99.0	96.7
Average text.	95.1	76.5	82.2	98.9	99.1	99.0	**99.9**	93.5	97.0	97.5
Bottle	99.9	89.2	94.2	**100**	99.2	**100**	**100**	89.2	98.4	97.4
Cable	81.9	75.7	83.2	91.8	70.3	99.5	**100**	81.7	52.7	90.1
Capsule	88.4	73.2	68.1	86.5	98.5	98.1	**100**	91.2	89.0	69.7
Hazelnut	83.3	78.5	85.5	95.7	**100**	**100**	**100**	97.8	84.5	94.5
Metal nut	88.5	70.0	66.7	96.9	98.7	**100**	**100**	90.4	92.8	96.0
Pill	83.8	74.3	78.6	90.2	98.9	**99.6**	**99.6**	80.6	80.9	95.0
Screw	84.5	74.6	**100**	95.7	93.9	98.1	97.8	99.9	20.3	61.0
Toothbrush	**100**	65.3	**100**	**100**	**100**	**100**	94.4	98.9	86.4	97.8
Transistor	90.9	79.2	84.3	95.8	93.1	**100**	98.8	93.0	65.0	91.6
Zipper	98.1	74.5	87.6	99.4	**100**	99.4	99.5	99.9	98.2	96.2
Average obj.	89.9	75.4	84.8	93.0	97.4	99.2	**99.5**	93.3	76.8	88.9
Average	91.7	76.2	83.9	95.0	98.0	99.1	**99.4**	93.0	83.5	90.4

Table 1: Results in anomaly detection (I-AUROC) on MVTec AD for various state-of-the-art approaches in comparison with ours.

4.2 Simplex noise

In the context of two-dimensional noise generation, Simplex noise is generated using a specific procedure. Initially, random gradients are sampled on a simplex grid consisting of equilateral triangles. When evaluating a candidate point, the inner product between the gradient and the offset of the candidate from the nearest three triangle vertices is computed. These resulting values are then interpolated to produce smooth noise. Intuitively, Simplex noise offers potential advantages over standard Gaussian perturbations by introducing more structured corruption. Consequently, denoising processes can effectively "repair" these structured anomalies, leveraging the inherent organization present in Simplex noise.

4.3 Gaussian Mixture Model

In this distribution noise is not sampled independently for each pixel, but instead sampled once for the whole image. Every noise map is made using function f that is created from K randomly sampled 2D Gaussian components. Each 2D Gaussian component $g_i(x, y)$ is defined as:

$$g_i(x,y) = \frac{1}{2\pi\sigma_{x_i}\sigma_{y_i}} \cdot e^{-\frac{(x-\mu_{x_i})^2}{2\sigma_{x_i}^2} - \frac{(y-\mu_{y_i})^2}{2\sigma_{y_i}^2}}, \quad (7)$$

where $g_i(x, y)$ is the i-th 2D Gaussian component of the function, μ_{xi} and μ_{yi} are the means of the i-th Gaussian component along the x and y dimensions, respectively, σ_{xi} and σ_{yi} are the standard deviations of the i-th Gaussian component along the x and y dimensions, respectively.

The overall 2D function $f(x, y)$ is a sum of all K 2D Gaussian components:

$$f(x,y) = \sum_{i=1}^{K} w_i \cdot g_i(x,y), \quad (8)$$

where w_i is the weight for the i-th component. w_i is sampled uniformly from $[0, 1]$. f is evaluated on $x \in$ $\{0, 1, ..., N-1\}, y \in \{0, 1, ..., N-1\}$, where N is equal to the size of the image.

5 Experiment

5.1 Dataset

Our research extensively relies on a widely recognized datasets for anomaly detection and localization: the MVTec Anomaly Detection Dataset [1]. The MVTec AD Dataset contains 6,612 images encompassing 5 texture categories and 10 object categories. To ensure consistency in our experiments, a standard preprocessing approach is employed. Each image is resized to dimensions of 256×256 and subsequently center-cropped to 224×224. Pixel-level annotations are provided for the test images, enabling accurate evaluation and analysis.

5.2 Evaluation metrics

The performance of anomaly detection at the image level is evaluated by employing the Area Under the Receiver Operator Curve (I-AUROC). For anomaly localization the pixel-wise AUROC (P-AUROC) and average precision (AP) are employed as evaluation metrics. These metrics allow us to accurately assess the effectiveness of our anomaly detection and localization algorithms.

5.3 Implementation details

This section describes the configuration implementation details of the experiments in this paper. For the diffusion process 1000 steps are used with a linear schedule ranging from 10^{-4} to 10^{-2}. ResUNet [33] is used for the base architecture of the model with the base number of channels set to 64. The model was trained for 2000 epochs using the AdamW optimizer with a batch size of 6. The learning rate was set to 10^{-4} and was multiplied by 0.1 after 1500 epochs. Each image was first diffused for 200 steps. For obtaining the reconstruction from the diffused image 200 reverse steps were performed. After obtaining the final mask, it was post-processed with a mean filter of

Method	Recon. Methods			State-of-the-art			Ours		
	[34]	[20]	[36]	[3]	[8]	[9]	Norm.	Simpl.	Gauss Mix. Model
Carpet	96.3 / **61.4**	87.0 / -	99.2 / -	95.5 / 53.5	99.0 / -	**99.4** / -	89.8 / 60.3	71.2 / 26.7	90.0 / 58.8
Grid	98. 8/ 36.4	94.0 / -	98.8 / -	**99.7** / 65.7	98.7 / -	98.3 / -	97.8 / **74.7**	92.3 / 43.5	88.9 / 38.1
Leather	95.7 / 51.6	78.0 / -	**99.5** / -	98.6 / 75.3	99.3 / -	**99.5** / -	98.6 / **84.5**	90.0 / 55.9	97.6 / 64.8
Tile	89.1 / 52.6	59.0 / -	94.4 / -	**99.2** / **92.3**	95.6 / -	96.3 / -	92.4 / 77.4	86.8 / 61.6	87.7 / 72.3
Wood	85.8 / 38.2	73.0 / -	88.7 / -	96.4 / **77.7**	95.0 / -	**97.0** / -	87.0 / 57.4	80.3 / 48.9	90.1 / 71.2
Average text.	93.9 / 48.0	78.2.0 / -	96.1 / -	**97.9** / **72.9**	97.5 / -	98.1 / -	93.9 / 74.5	86.1 / 47.3	90.9 / 60.9
Bottle	98.4 / 76.4	93.0 / -	97.1 / -	**99.1** / **86.5**	98.6 / -	97.7 / -	90.3 / 64.9	88.5 / 60.7	90.2 / 58.4
Cable	84.2 / 24.2	82.0 / -	91.0 / -	94.7 / **52.4**	**98.4** / -	**98.4** / -	85.9 / 39.6	49.5 / 10.0	83.2 / 35.7
Capsule	92.8 / 38.2	94.0 / -	97.7 / -	94.3 / 49.4	**98.8** / -	99.1 / -	96.2 / **65.4**	91.3 / 51.6	85.5 / 24.2
Hazelnut	96.1 / 33.8	97.0 / -	98.3 / -	**99.7** / **92.9**	98.7 / -	99.1 / -	97.4 / 84.5	90.8 / 50.7	92.3 / 65.9
Metal nut	92.5 / 64.3	89.0 / -	93.3 / -	**99.5** / **96.3**	98.4 / -	98.5 / -	88.5 / 62.4	80.5 / 59.1	87.6 / 76.9
Pill	95.7 / **51.6**	91.0 / -	98.3 / -	97.6 / 48.4	97.4 / -	**99.2** / -	81.8 / 40.9	89.3 / 43.9	88.6 / 48.9
Screw	**98.8** / 43.9	96.0 / -	99.5 / -	97.7 / **58.2**	99.4 / -	99.4 / -	98.5 / 70.3	83.6 / 11.0	86.4 / 17.0
Toothbrush	**98.9** / 50.6	92.0 / -	**98.9** / -	98.1 / 44.7	98.7 / -	**98.9** / -	94.1 / **63.9**	82.0 / 25.6	90.4 / 44.6
Transistor	87.7 / 39.2	80.0 / -	96.1 / -	90.9 / 50.7	96.3 / -	**97.3** / -	84.8 / **50.7**	71.8 / 23.3	76.8 / 41.2
Zipper	97.8 / 63.4	88.0 / -	**99.2** / -	98.8 / **81.5**	98.8 / -	98.7 / -	97.1 / 78.2	91.5 / 53.6	90.4 / 49.3
Average obj.	94.3 / 48.7	90.2 / -	96.9 / -	97.0 / **66.1**	98.4 / -	**98.6** / -	89.2 / 63.5	82.8 / 39.0	87.1 / 46.2
Average	94.2 / 48.2	86.2 / -	96.6 / -	97.3 / **68.4**	98.1 / -	**98.5** / -	92.3 / 66.2	83.1 / 42.3	88.4 / 51.1

Table 2: Results in anomaly localization (P-AUROC / AP) on MVTec AD for various state-of-the-art approaches in comparison with ours.

size 11×11 for the Gaussian and the Simplex noise distribution, while for Gaussian Mixture Model distribution a filter of size 5×5 was used. Filter sizes were determined empirically.

5.4 Results

Anomaly detection results are shown in Table 1. With Gaussian noise distribution we achieve an I-AUROC of 90.4%, with Simplex noise distribution 83.5% and with Gaussian Mixture Model noise distribution an I-AUROC of 93.0%. For anomaly localization the results are shown in Table 2. We achieve a P-AUROC of 92.3% and AP of 66.2% with the Gaussian noise distribution. With Simplex noise distribution we achieve 83.1% and 42.3% and with Gauss Mixture noise distribution we achieve 88.4% and 51.1%. Some sample results can be seen on Figure 2.
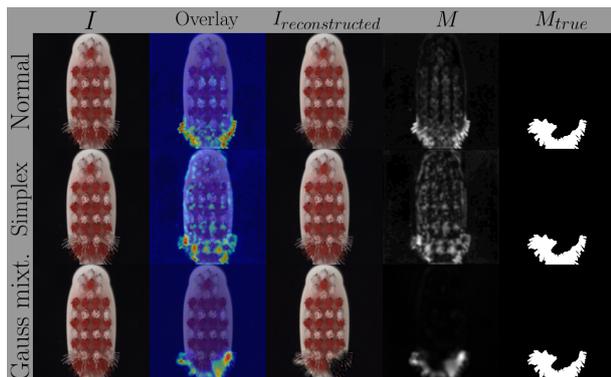

Figure 2: Sample results for all three noise distributions.

Gaussian noise distribution outperforms most of the previously proposed purely reconstruction methods. Previous reconstruction methods are mostly based upon auto-encoders (AE) [20], UNet [34] and GANs [14], which aren't currently as capable as diffusion models in the image generation aspect. It is most likely, that the diffu-

Number of steps	10	50	200	500
Normal	87.2	90.8	93.0	90.1
Simplex	81.7	82.4	83.5	82.8
Gauss Mix. Model	89.0	89.1	90.4	89.6

Table 3: Anomaly detection results (I-AUROC) for proposed noise distributions with different amount of reverse steps.

sion model does not over generalize as much as the previous methods and creates less false positives than previous methods. The Gaussian noise distribution lacks behind state-of-the-art methods, due to the fact that those methods are built upon different principles that are not purely reconstructive.

Results from the Gaussian noise distribution and the Gaussian Mixture Model noise distribution are comparable in most of the categories with two notable exceptions, the Cable and Screw category. In both of these categories there was an above average number of false positives. With the Screw category, we hypothesize the model did not learn the normality of the object due to the rotation of the objects. In the Cable category the model struggled to learn the normality of the copper wires, as it reconstructed them often, even though they weren't anomalous. The Simplex noise distribution had the problems with the false positives more often. We hypothesize the model learned too tight of a boundary for what is a normal object and what is not, resulting in an increase of false positives.

5.5 Ablation study

Number of reverse steps We explored the importance in the number of reverse steps for each of the noise distributions. The results can be seen in Table 3. As we can see there exists a finespot for each of the models which is not the same for each distribution.

Filter size	1	5	11	17
Normal	80.8	90.3	93.0	92.8
Simplex	80.3	81.5	83.5	82.8
Gauss Mix. Model	89.8	90.4	90.2	89.9

Table 4: Anomaly detection results (I-AUROC) for proposed noise distributions for different kernel sizes for mask postprocessing.

Size of the mean filter   In most anomaly detection approaches use a mean filter to smooth the final mask. And the size of the mean filter plays of a great importance in the final performance of our model. The results with different filter sizes can be seen in Table 4.

6   Conclusion

In this study, we investigated the effectiveness of different noise distributions, namely the standard Gaussian noise, Simplex noise, and a Gaussian Mixture Model noise, in the diffusion model for Surface Anomaly Detection. Our analysis revealed an improved performance of the Gaussian noise distribution over most of the previously proposed reconstruction methods with an I-AUROC of 93.0%. With the Gaussian Mixture Model noise distribution we achieved 90.4% I-AUROC, and with the Simplex noise distribution we achieved 81.5% I-AUROC. To gain further insights into the distribution of actual defects, our future work will involve exploring more complex noise distributions, aiming to enhance our understanding and detection capabilities in surface anomaly detection.

7   Acknowledgements

References

[1] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9592–9600, 2019.

[2] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "SPot-the-Difference self-supervised pre-training for anomaly detection and segmentation," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXX*, pp. 392–408, Springer, 2022.

[3] V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM-a discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8330–8339, 2021.

[4] M. Yang, P. Wu, and H. Feng, "MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities," *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105835, 2023.

[5] V. Zavrtanik, M. Kristan, and D. Skočaj, "DSR–a dual subspace re-projection network for surface anomaly detection," in *European Conference on Computer Vision*, pp. 539–554, Springer, 2022.

[6] J. Bae, J.-H. Lee, and S. Kim, "Image anomaly detection and localization with position and neighborhood information," *arXiv preprint arXiv:2211.12634*, 2022.

[7] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "SimpleNet: A simple network for image anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402–20411, 2023.

[8] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2022.

[9] J. Yu, Y. Zheng, X. Wang, W. Li, Y. Wu, R. Zhao, and L. Wu, "Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows," *arXiv preprint arXiv:2111.07677*, 2021.

[10] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Fully convolutional cross-scale-flows for image-based defect detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1088–1097, 2022.

[11] M. Tailanian, Á. Pardo, and P. Musé, "U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold," *arXiv preprint arXiv:2211.12353*, 2022.

[12] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, pp. 4–11, 2014.

[13] D. T. Nguyen, Z. Lou, M. Klar, and T. Brox, "Anomaly detection with multiple-hypotheses predictions," in *International Conference on Machine Learning*, pp. 4800–4809, PMLR, 2019.

[14] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Asian conference on computer vision*, pp. 622–637, Springer, 2018.

[15] Y. Liang, J. Zhang, S. Zhao, R. Wu, Y. Liu, and S. Pan, "Omni-frequency channel-selection representations for unsupervised anomaly detection," *arXiv preprint arXiv:2203.00259*, 2022.

[16] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[17] H. Zhang, Z. Wang, Z. Wu, and Y.-G. Jiang, "DiffusionAD: Denoising diffusion for anomaly detection," *arXiv preprint arXiv:2303.08730*, 2023.

[18] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 650–656, June 2022.

[19] A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein, "Cold diffusion: Inverting arbitrary image transforms without noise," *arXiv preprint arXiv:2208.09392*, 2022.

[20] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," pp. 372–380, 01 2019.

[21] K. Batzner, L. Heckler, and R. König, "EfficientAD: Accurate visual anomaly detection at millisecond-level latencies," *arXiv preprint arXiv:2303.14535*, 2023.

[22] J. Jang, E. Hwang, and S.-H. Park, "N-Pad: Neighboring pixel-based industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 4364–4373, June 2023.

[23] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "DiffWave: A versatile diffusion model for audio synthesis," in *International Conference on Learning Representations*, 2021.

[24] X. L. Li, J. Thickstun, I. Gulrajani, P. Liang, and T. Hashimoto, "Diffusion-LM improves controllable text generation," in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, eds.), 2022.

[25] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. van den Berg, "Structured denoising diffusion models in discrete state-spaces," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17981–17993, 2021.

[26] R. Huang, M. W. Lam, J. Wang, D. Su, D. Yu, Y. Ren, and Z. Zhao, "FastDiff: A fast conditional diffusion model for high-quality speech synthesis," 2022.

[27] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

[28] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8780–8794, 2021.

[29] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, pp. 8162–8171, PMLR, 2021.

[30] Y. Benny and L. Wolf, "Dynamic dual-output diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11482–11491, 2022.

[31] E. Nachmani, R. S. Roman, and L. Wolf, "Non gaussian denoising diffusion models," *arXiv preprint arXiv:2106.07582*, 2021.

[32] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*, pp. 2256–2265, PMLR, 2015.

[33] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.

[34] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, p. 107706, 2021.

[35] Y. Fei, C. Huang, J. Cao, M. Li, Y. Zhang, and C. Lu, "Attribute restoration framework for anomaly detection," *IEEE Transactions on Multimedia*, vol. PP, pp. 1–1, 12 2020.

[36] J. Pirnay and K. Chai, "Inpainting transformer for anomaly detection," in *Image Analysis and Processing – ICIAP 2022* (S. Sclaroff, C. Distante, M. Leo, G. M. Farinella, and F. Tombari, eds.), (Cham), pp. 394–406, Springer International Publishing, 2022.