jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2023, VOL. 69, NO. 4, PP. 719-726
Manuscript received August 21, 2023; revised October, 2023. DOI: 10.24425/ijet.2023.147693

ADDP: Anomaly Detection Based on Denoising Pretraining

Xianlei Ge, Xiaoyan Li, and Zhipeng Zhang

Abstract—**Acquiring labels in anomaly detection tasks is expensive and challenging. Therefore, as an effective way to improve efficiency, pretraining is widely used in anomaly detection models， which enriches the model's representation capabilities, thereby enhancing both performance and efficiency in anomaly detection. In most pretraining methods, the decoder is typically randomly initialized. Drawing inspiration from the diffusion model, this paper proposed to use denoising as a task to pretrain the decoder in anomaly detection, which is trained to reconstruct the original noise-free input. Denoising requires the model to learn the structure, patterns, and related features of the data, particularly when training samples are limited. This paper explored two approaches on anomaly detection: simultaneous denoising pretraining for encoder and decoder, denoising pretraining for only decoder. Experimental results demonstrate the effectiveness of this method on improving model's performance. Particularly, when the number of samples is limited, the improvement is more pronounced.**

Keywords—**Anomaly Detection; Diffusion Models; Image Denoising; Pretraining; Transfer Learning**

I. Introduction

ANOMALY detection in medical imaging aims to identify and locate abnormal regions in various types of medical images, including X-ray, CT scans, and MRI [1]. This technique has diverse applications in clinical practice. It can be employed for early detection, diagnosis, and localization of diseases, including lung nodules, tumor segmentation, and stroke recognition. The advancement of this technology equips doctors with precise, rapid, and reliable tools to enhance the management and treatment of patients' health is-sues. Simultaneously, this technology significantly assists doctors in early diagnosis, treatment planning, and disease progression monitoring.

The rapid advancement of computer and artificial intelligence technology has led to the widespread attention and adoption of neural networks for detecting abnormalities in medical images [2]. It plays a crucial role in various tasks, including disease screening, diagnosis, prediction of disease development trends, and detection of organ and tissue lesions.

However, the practical application of deep learning in medical anomaly detection poses several challenges, with the sample size problem being the most prominent. Due to numerous limitations, acquiring large-scale annotation data for medical images is prohibitively expensive, and annotating medical images necessitates extensive expertise and experience from medical professionals. Pretraining is a viable strategy to partially address the issue of limited sample size in medical image anomaly detection, which involves initially training the model on a large-scale dataset and subsequently fine-tuning it on the target task using the learned parameters as initial values [3]. In medical anomaly detection, pretraining can utilize large-scale non-medical image data or conventional medical data to acquire general feature representations. By transferring these learned features, the scarcity of medical image data in anomaly detection tasks can be mitigated.

Currently, in medical image anomaly detection, pretraining primarily focuses on encoders, using classification tasks as a guide, while disregarding decoders and initializing their parameters randomly. This approach can result in suboptimal models. In this paper, we proposed ADDP, which undergoes pretraining for denoising tasks to enable a series of initializations. Denoising pretraining can be considered a form of self-supervised learning, where the decoder is trained to reconstruct the original input without noise. This task necessitates the decoder to grasp the structure, patterns, and relevant features of the data in order to minimize the reconstruction error. Consequently, the decoder will acquire a representation that demonstrates robustness against noise and irregularities in the input data. Furthermore, there exist several additional advantages to employing denoising tasks for pretraining. Firstly, the denoising task is an unsupervised learning task that operates without label information, enabling the utilization of extensive unlabeled data for pretraining. Consequently, it compensates for the expensive and challenging label acquisition process. Secondly, pretraining with denoising tasks can yield stable feature representations, diminish noise interference in training samples, and enhance the model's generalization ability and robustness. The contributions of this paper can be summarized as follows:

(1) This work proposed a method for initializing parameters in anomaly detection guided by denoising, which effectively enhances the performance of anomaly detection.

This work was supported by the University Natural Science Foundation of Anhui Province (Grant No.2022AH051578,2023AH051551), Key Science Research Foundation of Huainan Normal University (Grant No.2022XJZD019) and Guiding Science and Technology Foundation of Huainan (Grant No.2020050).

Xianlei Ge is with School of Electronic Engineering, Huainan Normal University, China and College of Computing and Information Technologies, National University, Philippines (e-mail: gex@students.national-u.edu.ph).

Xiaoyan Li is with School of Computer, Huainan Normal University, China (e-mail: lix@students.national-u.edu.ph).

Zhipeng Zhang is with School of Electronic Engineering, Huainan Normal University, China (e-mail: 13135545637.zhang@gmail.com).

(2) This paper optimizes several techniques within the denoising process to maximize its benefits for the model.

(3) This work explored two denoising pretraining methods. One can choose to conduct denoising pretraining on both the encoder and decoder simultaneously, depending on the situation, or solely on the decoder.

(4) This work conducted extensive experiments to examine the influence of different denoising variables, and these findings can serve for multiple other tasks.

II. Related Work

A. Medical anomaly detection based on deep learning

Medical anomaly detection based on deep learning is a widely researched and significant problem in the field. One prominent aspect of deep learning is its capability to model non-linear relationships. By increasing the non-linearity in the model, it becomes possible to achieve improved separation between normal and abnormal samples, as well as better modeling of inconsistencies within the data. Deep learning methods can be classified into two main categories: unsupervised methods and supervised methods. Unsupervised methods commonly employ autoencoders (AE) [4] and generative adversarial networks (GAN) [5]. For instance, Lu et al. proposed a VAE framework for detecting skin image abnormalities [6], while Zimmerer et al. enhanced VAE for MRI anomaly detection [7]. Uzunova et al. also utilized VAE for pathological detection [8]. Notably, Schlegl et al. applied GAN to medical anomaly detection, representing significant contributions in this area [9]. On the other hand, supervised learning methods frequently rely on convolutional neural networks (CNN). Esteva et al. employed CNN for skin cancer detection [10], and Turner et al. utilized deep belief networks for detecting abnormal signals in electrocardiograms (ECG) [11]. Wang et al. employed cascaded anisotropic convolutional neural networks for segment brain tumor [12]. Overall, deep learning has emerged as a crucial technology in medical anomaly detection, with continual updates and iterations in training techniques and network models to enhance performance.

B. Diffusion models

Diffusion models have a long-standing history in the field of machine learning [13,14]. Their primary objective is to train models to eliminate noise from data and distinguish between noisy and clean data. In recent developments, the denoising diffusion probability model (DDPM) [15]has demonstrated unparalleled performance in the domains of image and text generation. It significantly outperforms alternative generative models in terms of both density estimation and sample quality, while also exhibiting extensive pattern coverage [16–19]. DDPM is a highly potent model that achieves its capabilities by learning to transform Gaussian noise into the target distribution via a series of iterative denoising steps, approximating complex empirical distributions [20]. Its performance is truly impressive. The concept of denoising can yield numerous additional advantages for contemporary neural network methodologies.

C. Pretraining

Pretraining is a crucial technique in deep learning, used to initialize the parameters of neural network models by initially training them on large-scale datasets. The objective of pretraining is to acquire a generalized feature representation for a given task and subsequently utilize these learned features to address specific downstream tasks. The pretraining method capitalizes on prior knowledge during the pretraining process, encompassing an understanding of the image's structure, texture, and semantics. Despite a reduction in the number of training sets, this prior knowledge remains valuable in assisting the model to improve its performance in image tasks. Pretraining is also a valuable approach in the domain of medical image processing, proven to enhance model performance [3,21,22]. Typically, pretraining involves training the encoder (backbone) either as a classifier [23] or as a self-supervised feature extractor [24–26]. Our observation reveals that while there exists a plethora of methods for pretraining the encoder, the decoder is frequently initialized randomly. Considering the decoder's significant role in tasks involving pixel-level output granularity, attention must be given to its pretraining.

III. Methodology

We proposed employing denoising pretraining for encoder and decoder (DPED), as well as denoising pretraining for decoder (DPD), to discover an improved pretraining approach. The former approach employs a one-stage process that directly utilizes the autoencoder within the data, while the latter employs a two-stage process. It first trains the encoder guided by other tasks, then freezes its parameters and trains the decoder guided by denoising. This chapter will provide a detailed introduction to both methods.

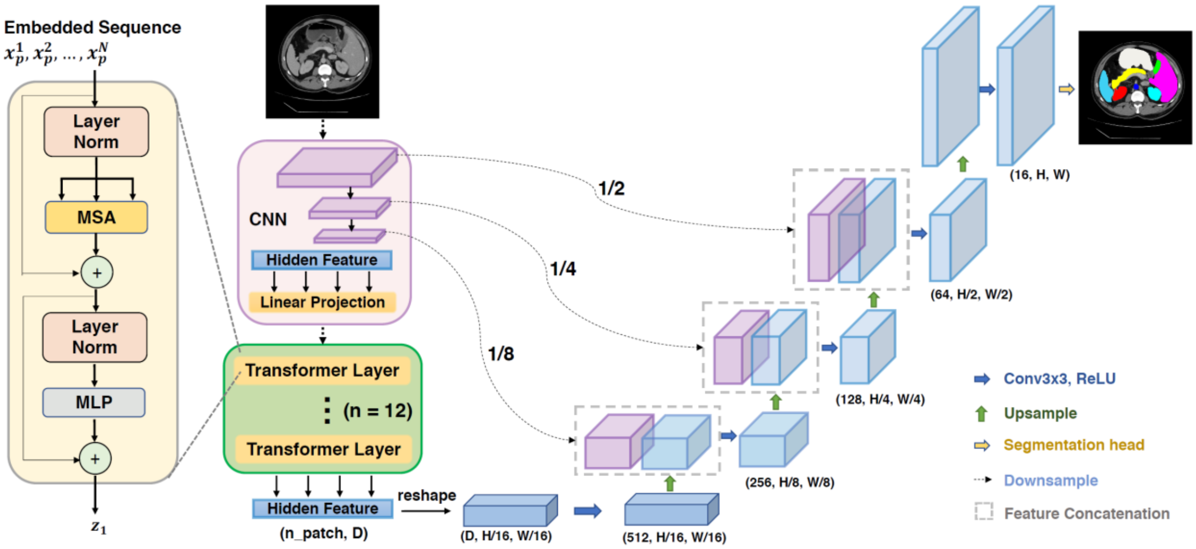

Fig.1. Trans UNet [27]

A. Encoder and decoder

This study aims to define an autoencoder denoted as f_θ, g_ϕ, which comprises two sets of parameters θ and ϕ, allowing for their transferability. The autoencoder takes $X \subset \mathbb{R}^{H\times W\times C}$ as input, where f_θ generates a high dimensional tensor, and g_ϕ produces mask. The implementation of Trans-UNet [27] is chosen for this paper, as depicted in Figure 1. UNet is a suitable choice for diffusion work and is regarded as the optimal model for mapping noise to the original distribution. Trans-UNet employs a hybrid model that combines CNN and Transformer. It initially employs CNN to generate feature maps and then uses Transformer for feature processing. This approach enables the utilization of deep features mentioned by CNN in the decoder, while encoding image features as a sequence provides significant global context information. This enhancement aids in comprehending the semantics and structure of the image, ultimately leading to improved segmentation performance.

B. Denoising objective

The objective of the denoising task is to remove noise from the data. There are various methods to accomplish this task, such as predicting clean data directly. For instance, given an image sample x, Gaussian noise ε can be added to create x_ϵ.

$$x_\epsilon = x + \sigma\varepsilon \tag{1}$$

Among them, $\varepsilon \sim N(0, I)$, σ corresponds to a fixed standard deviation. Equation (1) provides a straightforward method for introducing noise. Moreover, the direct addition of noise significantly influences the final outcome. Equation (1) has been enhanced to exert control over the impact of noise on the final result while preserving the original data information. This improvement is demonstrated in (2).

$$x_\sigma = \frac{1}{\sqrt{1+\sigma^2}}(x + \sigma\varepsilon) \tag{2}$$

Currently, it can control the influence of noise on the final result to a certain extent. When $\sigma = 1$, it corresponds to (1), while (2) is better suited for tasks that exhibit sensitivity to the data's norm. Let the encoder be denoted as f_θ and the decoder as g_ϕ. Consequently, the objective function can be formulated in (3):

$$L_1 = \left\| g_\phi\left(f_\theta(x_\sigma)\right) - x \right\|_2^2 \tag{3}$$

However, Ho et al. demonstrated that incorporating noise as the loss function during training the diffusion model yields better results. In this approach, the final prediction target is changed from x to ε. The objective function can be written as:

$$L_2 = \left\| g_\phi\left(f_\theta(x_\sigma)\right) - \varepsilon \right\|_2^2 \tag{4}$$

Methods targeting predictive noise can reduce prior assumptions regarding the input image, which assumes a specific distribution or structure. In contrast, the diffusion model can adapt to various types and levels of noise without prior knowledge of specific details in the input image, enhancing generalization and robustness. We have thoroughly investigated this choice in the subsequent experimental chapters.

C. Denoising pretraining for encoder and decoder

When denoising pretraining is performed simultaneously on the encoder and decoder, it can also be considered as the denoising pretraining of the complete model, as depicted in Figure 2. Let $\{X_1, \dots, X_N\} \subset \mathbb{R}^{H\times W\times C}$ represent an unlabeled dataset that can be normal samples or abnormal samples. Initially, noise is added to this dataset using (2), resulting in $\{X_{\epsilon 1}, \dots, X_{\epsilon N}\} \subset \mathbb{R}^{H\times W\times C}$. Subsequently, Equation (4) is utilized to train f_θ and g_ϕ. The parameters θ and ϕ are then fine-tuned on the target task dataset to obtain the optimal set.

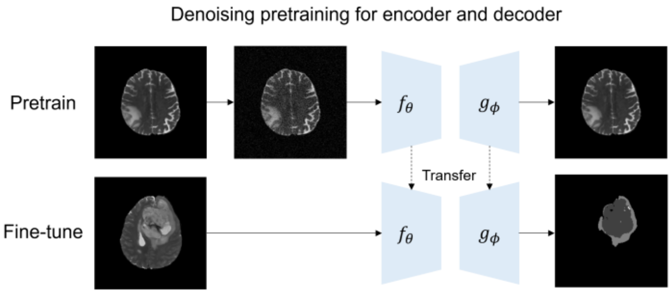

Fig.2. Denoising pretraining for encoder and decoder (DPED)

D. *Denoising pretraining for only decoder*

The process of pretraining only the decoder is illustrated in Figure 3, comprising three steps. Initially, the encoder is trained using alternative tasks. We suggest employing classification tasks to guide the encoder towards convergence. Let $\{X_1,\dots,X_N\} \subset \mathbb{R}^{H\times W\times 3}$ be a sample dataset and $\{Y_1,\dots,Y_n\}\mathbb{R}^{H\times W\times\{1,\dots,K\}}$ be its corresponding label set, where K represents the number of categories. The encoder f_θ is trained until convergence, after which the parameters θ are frozen. Subsequently, the decoder g_ϕ is trained using denoising training. Noise is added to the dataset $\{X_1,\dots,X_N\} \subset \mathbb{R}^{H\times W\times 3}$ to obtain $\{X_{\epsilon 1},\dots,X_{\epsilon N}\} \subset \mathbb{R}^{H\times W\times 3}$, and (4) is utilized for training to determine the optimal parameter ϕ. Finally, the parameters θ are unfrozen, and both f_θ and g_ϕ are fine-tuned using the target task dataset.

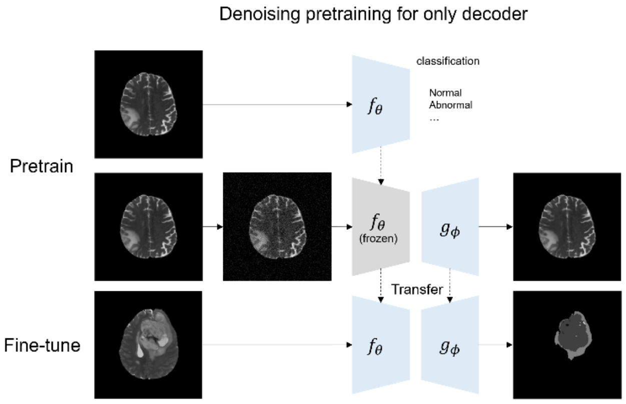

Fig.3. Denoising pretraining for only decoder (DPD)

E. *Noise level*

The noise level σ is a critical hyperparameter. Insufficiently small noise may hinder the model's ability to learn meaningful feature representations. While excessive noise will cause a significant shift between the distribution of the noise image and the original image. Consequently, the model becomes incapable of effectively learning and capturing the key features of the original image. Thus, this greatly impacts the model's overall performance and effectiveness. Figure 4 illustrates examples of varying noise levels.

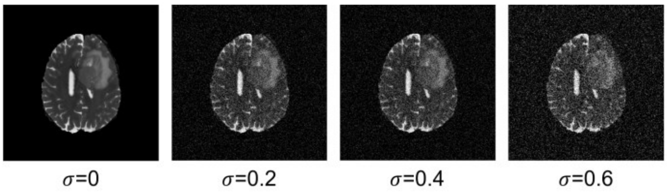

Fig.4. Different noise levels

We set σ to 0.4, and this value will be further examined in the subsequent experimental sections.

IV. Experiment Results and Analysis

This chapter aims to conduct a series of experiments to investigate the impact of denoising pretraining on anomaly detection.

A. *Experiment environment*

Table I presents the detailed configuration of hardware and software utilized during the experimental phase.

Table I
Experiment environment

Environment	Configuration
CPU	Intel Core i7 12700k
GPU	Nvidia Tesla V100 32GB
Memory	DDR4 16GB
Hard Disc	WestData SSD 1TB
Operating System	Windows11
Python	3.8
torch	1.12.1

B. *Datasets*

BraTS (Brain Tumor Segmentation) is an annual challenge and dataset focused on brain tumor segmentation and classification. The challenge provides a platform for researchers and practitioners to evaluate and compare their algorithms for brain tumor analysis. The BraTS 2021 dataset consists of multimodal magnetic resonance imaging (MRI) scans of the brain, including T1-weighted (T1), T1-weighted with contrast enhancement (T1Gd), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) images. These different MRI sequences capture distinct aspects of brain tissue and provide complementary information for tumor analysis [28].

We utilized the data from Task 1: Brain Tumor Sub-Region Segmentation. The dataset provided for this task comprises 8,000 MRI scans obtained from 2,000 glioma patients.

C. *Metrics*

Dice coefficient, also known as Dice similarity coefficient or F1 score, is a common metric used to evaluate the similarity or overlap between two sets, particularly in the context of segmentation tasks. It is widely used in medical image analysis to assess the accuracy of segmentation algorithms. The Dice coefficient ranges from 0 to 1, where a value of 1 indicates perfect overlap or similarity between the two sets. The calculation formula for the Dice coefficient, shown in (5), represents the predicted segmentation as set A and the actual segmentation as set B.

$$\text{Dice coefficient} = \frac{2\times|A\cap B|}{|A|+|B|} \tag{5}$$

D. *Results analysis*

Table II presents the results obtained from four distinct approaches: without pretraining, encoder pretraining, denoising pretraining for encoder and decoder (DPED), and denoising pretraining for only decoder (DPD). Without pretraining, both encoder and decoder are randomly initialized. Encoder pretraining, on the other hand, involves pretrain the encoder for classification tasks, while the decoder remains randomly initialized.

TABLE II
THE RESULTS OBTAINED FROM THE THREE METHODS.

Method [a]	100%	50%	10%	5%	1%
Without Pretraining	0.59	0.43	0.31	0.23	0.19
Encoder Pretraining	0.85	0.8	0.73	0.69	0.66
DPED	0.87	0.84	0.79	0.75	0.74
DPD	0.88	0.84	0.8	0.73	0.72

[a] The first row represents the proportion of the subset

Denoising pretraining methods exhibit superior performance. When the number of training samples significantly decreases, the results without pretraining become nearly unattainable. Conversely, the two denoising pretraining methods consistently achieve significant performance enhancements. Compared to solely pretrain the encoder, the benefits of denoising pretraining progressively increase as the number of samples decreases. The utilization of denoising pretraining enables the provision of a more comprehensive supervision signal by establishing the corresponding relationship between the noisy and clean images. This facilitates the model's capacity to effectively learn valuable information within the image.

In situations where there are an ample number of training samples, DPD yields superior performance. The encoder's role involves extracting high level semantic features from images, while the decoder is responsible for converting these features into pixel level predictions. In cases where only the decoder undergoes denoising pretraining, the model primarily focuses on learning and optimizing the decoder component, particularly during pixel-level prediction, to reconstruct the image details accurately.

Conversely, when the training samples are limited, DPED enables comprehensive learning of both semantic and detailed image information. Through cooperative training between the encoder and decoder, this approach enhances the model's expressive and generalization capabilities. Consequently, in scenarios with limited training samples, this method effectively utilizes available data to construct a more robust representation, thereby improving the performance of image segmentation.

E. Ablation studies

In the preceding chapters, we developed a range of techniques aimed at enhancing the effectiveness of pretraining. These techniques include optimizing the model's structure, modifying the process of transforming and introducing noise, converting predicted images into predicted noise, and selecting appropriate levels of noise. This section includes a series of experiments designed to investigate the impact of the aforementioned methods.

1) Architecture of encoder and decoder

We selected Trans-UNet as our model for this study. To investigate the impact of the denoising pretraining method proposed in this paper on various models, this section includes other models including UNet [29], Attention-UNet [30] and SegNet [31]. We initialized different architectures with three variations: without pretraining, DPED, and DPD. In the case of DPD, we added three linear layers to the output layer in order to output categories for the second step of pretraining. As presented in Table 3, the other three models exhibited significant performance improvements when subjected to denoising pretraining. This demonstrates the universal and effective nature of the proposed denoising pretraining method across different models. Therefore, we consider it an outstanding pretraining technique.

TABLE III
THE RESULTS OF THREE DIFFERENT METHODS ON DIFFERENT MODELS.

Model	Method	100%	50%	10%	5%	1%
UNet	Without Pretraining	0.52	0.39	0.26	0.14	0.1
	DPED	0.79	0.77	0.75	0.72	0.69
	DPD	0.82	0.8	0.71	0.68	0.67
Attention UNet	Without Pretraining	0.56	0.42	0.28	0.17	0.14
	DPED	0.83	0.83	0.77	0.73	0.7
	DPD	0.85	0.81	0.76	0.7	0.69
SegNet	Without Pretraining	0.49	0.3	0.25	0.16	0.13
	DPED	0.82	0.79	0.75	0.71	0.68
	DPD	0.81	0.77	0.72	0.69	0.65

2) *Noise Schedule*

We have implemented improvements in the method of introducing noise, transitioning from (1) to (2), in order to regulate the impact of noise on the final outcome. This section investigates the impact of these enhancements, as demonstrated in Table 4. By utilizing (2) to introduce noise, we observe a discernible improvement. This enhancement stems from the effective regulation of noise within (2), allowing the model to acquire a more profound comprehension of the correlation between noise and image [33], consequently enhancing its ability to handle and comprehend noise. Simultaneously, employing (2) enables a more precise adjustment of image distribution and diminishes the disparity between the distributions of clean and noisy images. Consequently, this refinement enhances the transferability of pretrained representations to the ultimate task.

TABLE IV
CORRESPONDING RESULTS OF DIFFERENT LOSS FUNCTIONS.

Method	Predict	100%	50%	10%	5%	1%
DPED	x	0.82	0.8	0.76	0.73	0.68
	ε	0.87	0.84	0.79	0.75	0.74
DPD	x	0.84	0.81	0.71	0.66	0.6
	ε	0.88	0.84	0.8	0.73	0.72

3) *Noise level*

We investigated the impact of various noise levels on pretraining [34], recognizing noise level as a critical hyperparameter. Figure 5 illustrates that lower levels of noise do not contribute positively to pretraining, whereas excessive noise negatively affects the model's performance. The DPED, whose encoder and decoder are pretrained by denoising, is more susceptible to noise, whereas the DPD, which only pretrains the decoder through denoising, exhibits greater stability. Excessive noise can also impact the pretraining process, leading to numerous negative effects by excessive focus on noise. Therefore, we chose a noise level of 0.4 to strike a balance between sufficient learning signals and minimizing the impact on model accuracy.

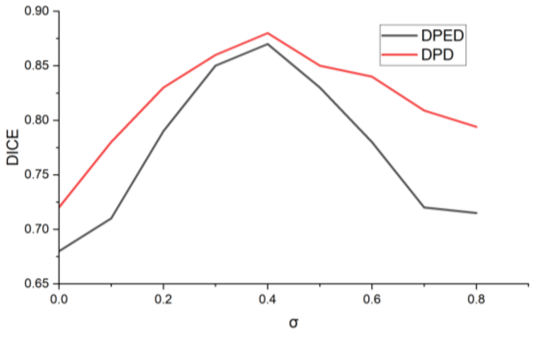

Fig.5. Different noise level

F. *Discussion*

Denoising pretraining methods are an effective approach for enhancing the feature learning capability of models, which can lead to improved performance in segmenting abnormal regions. By training a model to remove noise and restore signals, the model develops a better understanding of the underlying structure and important information within the data. In the denoising pretraining process, the model is exposed to various types and levels of noise intentionally added to the input data. This helps the model learn to distinguish between the noise and the essential signal, allowing it to focus on the relevant features and discard the irrelevant ones. This selective attention helps the model identify and emphasize the discriminative aspects of the data, which are crucial for accurately segmenting abnormal regions. As the model becomes proficient in noise removal and signal restoration, it becomes more adept at capturing subtle patterns and details in the data. This acquired skill enables the model to recognize important image structures, such as edges, textures, and shapes, even in the presence of noise or other types of interference. By extracting robust and discriminative features, the model gains a deeper understanding of the underlying characteristics of abnormal regions, leading to improved segmentation performance. Furthermore, the denoising pretraining process encourages the model to learn more generalizable representations. By exposing the model to diverse noise patterns, it becomes more robust against different types of variations and artifacts that can be present in real-world data. This robustness helps the model perform well on new and unseen data, enhancing its ability to accurately segment abnormal regions across different datasets and clinical scenarios. In summary, denoising pretraining methods significantly contribute to feature learning by enhancing a model's capacity to learn discriminative and robust representations. The acquired skill of noise removal and signal restoration enables the model to focus on crucial information and image structure, leading to improved performance in segmenting abnormal regions. By learning to extract discriminative features and developing robustness against various noise patterns, the model becomes more effective in accurately identifying and delineating abnormal regions in medical images.

V. CONCLUSION

This paper proposed a pretraining method called ADDP for anomaly detection in medical images. The method involves dual pretraining approaches: Denoising pretraining for encoder and decoder (DPED) and Denoising pretraining for only decoder (DPD). The parameters are initialized based on the denoising task. Simultaneously, the denoising task is optimized to enhance the model's ability to learn valuable feature representations during the pretraining process. Several experiments have demonstrated that denoising guided pretraining can effectively enhance model performance. Specifically, in scenarios with limited samples, the pretrained denoising model exhibits greater stability. Moreover, incorporating multiple enhancements to the denoising tasks significantly boosts model performance. The proposed denoising pretraining method presented in this paper is not only applicable to medical image anomaly detection but also readily transferable to other tasks.

REFERENCES

[1] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep Learning for Medical Anomaly Detection – A Survey," ACM Comput. Surv., vol. 54, no. 7, Jul. 2021. https://doi.org/10.1145/3464423

[2] P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial Intelligence in Medical Diagnosis," Ann. Intern. Med., vol. 108, no. 1, pp. 80–87, 1988. https://doi.org/10.7326/0003-4819-108-1-80

[3] Y. Qiu, F. Lin, W. Chen, and M. Xu, "Pre-training in Medical Data: A Survey," Machine Intelligence Research, vol. 20, no. 2, pp. 147-179, Apr. 2023. https://doi.org/10.1007/s11633-022-1382-8

[4] G. E. Hinton and R. S. Zemel, "Autoencoders, Minimum Description Length and Helmholtz Free Energy," in Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93, Denver, Colorado, pp. 3-10. 1993

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," Commun. ACM, vol. 63, no. 11, pp. 139-144, Nov. 2020. https://doi.org/10.1145/3422622

[6] Lu, Yuchen, and Peng Xu. "Anomaly detection for skin disease images using variational autoencoder." arXiv preprint arXiv:1807.01349 (2018). https://doi.org/10.48550/arXiv.1807.01349

[7] Zimmerer, David, et al. "Context-encoding variational autoencoder for unsupervised anomaly detection." arXiv preprint arXiv:1812.05941 (2018). https://doi.org/10.48550/arXiv.1812.05941

[8] H. Uzunova, S. Schultz, H. Handels, and J. Ehrhardt, "Unsupervised Pathology Detection in Medical Images Using Conditional Variational Autoencoders," International Journal of Computer Assisted Radiology and Surgery, vol. 14, no. 3, pp. 451-461, Mar. 2019. https://doi.org/10.1007/s11548-018-1898-0

[9] Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., & Schmidt-Erfurth, U. (2019). f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis, 54, 30-44. https://doi.org/10.1016/j.media.2019.01.010

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, Feb. 2017. https://doi.org/10.1038/nature21056

[11] J. Turner, A. Page, T. Mohsenin, and T. Oates, "Deep Belief Networks used on High Resolution Multichannel Electroencephalography Data for Seizure Detection," arXiv:1708.08430, 2017. https://arxiv.org/abs/1708.08430

[12] G. Wang, W. Li, S. Ourselin, T. Vercauteren, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2018), vol. 10670 LNCS, pp. 178-190. https://doi.org/10.1007/978-3-319-75238-9_16

[13] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol. "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103. 2008. https://doi.org/10.1145/1390156.1390294

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," Journal of Machine Learning Research, vol. 11, no. 110, pp. 3371-3408, 2010. https://dl.acm.org/doi/10.5555/1756006.1953039

[15] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33,6840-6851,2020.

[16] S. Bond-Taylor, A. Leach, Y. Long and C. G. Willcocks, "Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7327-7347, 1 Nov. 2022. https://doi.org/10.1109/TPAMI.2021.3116668

[17] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," in Proc. Advances in Neural Information Processing Systems, pp. 8780-8794, Curran Associates, Inc., 2021.

[18] D. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational Diffusion Models," in Advances in Neural Information Processing Systems, pp. 21696-21707, Curran Associates, Inc., 2021.

[19] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion Models: A Comprehensive Survey of Methods and Applications," arXiv preprint arXiv:2209.00796, 2023. https://doi.org/10.48550/arXiv.2209.00796

[20] J. Wyatt, A. Leach, S. M. Schmon and C. G. Willcocks, "AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 2022, pp. 649-655, https://doi.org/10.1109/CVPRW56347.2022.00080

[21] L. Zhou, H. Liu, J. Bae, J. He, D. Samaras, and P. Prasanna, "Self Pre-training with Masked Autoencoders for Medical Image Classification and Segmentation," arXiv preprint arXiv:2203.05573, 2023. https://doi.org/10.48550/arXiv.2203.05573

[22] Y. Tang et al., "Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 20698-20708, https://doi.org/10.1109/CVPR52688.2022.02007

[23] E. H. Eldeeb, A. M. Nagah, I. A. S. Amin, H. Kamel, and S. Fouad, "A Robust CNN Model for Diagnosis of COVID-19 Based on CT Scan Images and DL Techniques," in International Journal of Electronics and Telecommunications, vol. 68, no. 4, pp. 731–739, 2022. DOI: 10.24425/ijet.2022.143879. https://doi.org/10.24425/ijet.2022.143879

[24] A. van den Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," arXiv preprint arXiv:1807.03748, 2019. https://doi.org/10.48550/arXiv.1807.03748

[25] R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio, "Learning Deep Representations by Mutual Information Estimation and Maximization," 2019. https://doi.org/10.48550/arXiv.1808.06670

[26] P. Bachman, R. D. Hjelm, and W. Buchwalter, "Learning Representations by Maximizing Mutual Information Across Views," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2019.

[27] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arXiv preprint arXiv:2102.04306, 2021. https://doi.org/10.48550/arXiv.2102.04306

[28] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," arXiv preprint arXiv:2107.02314, 2021.
https://doi.org/10.48550/arXiv.2107.02314

[29] Ronneberger, O., Fischer, P. and Brox, T. Ronneberger. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351:234–41. 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[30] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," eprint arXiv:1804.03999, 2018.
https://doi.org/10.48550/arXiv.1804.03999

[31] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017.
https://doi.org/10.1109/TPAMI.2016.2644615

[32] J. Shang, T. Ma, C. Xiao, and J. Sun, "Pre-training of Graph Augmented Transformers for Medication Recommendation," arXiv preprint arXiv:1906.00346, 2019. https://doi.org/10.48550/arXiv.1906.00346

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words:

Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2021. https://doi.org/10.48550/arXiv.2010.11929

[34] T. Reiss, N. Cohen, L. Bergman and Y. Hoshen, "PANDA: Adapting Pretrained Features for Anomaly Detection and Segmentation," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 2805-2813, 2021. https://doi.org/10.1109/CVPR46437.2021.00283