InvDetect: Unsupervised Medical Anomaly Detection in the Noise Latent Space of DDIM

Xinyu Ma[1], Jinhui Ma[1], Shiqi He[2], Jian Pei[3], and Lingyang Chu[1⋆]

[1] McMaster University, Hamilton, ON, Canada
{ma209,maj26,chul9}@mcmaster.ca
[2] University of Michigan, Ann Arbor, MI, USA
shiqihe@umich.edu
[3] Duke University, Durham, NC, USA
j.pei@duke.edu

Abstract. Anomaly detection in medical images is useful for early diagnosis and treatment planning. Unsupervised methods are attractive because they avoid manual annotations and generalize well to unseen anomalies. However, existing unsupervised methods face fundamental limitations: reconstruction-based methods often suffer from pixel-level mismatch of generated pseudo-normal images, and embedding-based methods often fail to separate normal and abnormal images sufficiently in embedding spaces. In this paper, we propose an unsupervised method named *InvDetect* to address both limitations. InvDetect performs anomaly detection in a noise latent space induced by Denoising Diffusion Implicit Model (DDIM) inversion, which avoids generating pseudo-normal images thus prevents pixel-level mismatch. By training DDIM exclusively on normal image patches, InvDetect constructs a structured noise latent space in which normal patches form a compact cluster and abnormal patches deviate significantly. This yields better separation between normal and abnormal patches to produce a more reliable pixel-wise anomaly map. At last, we refine the anomaly map by enforcing spatial contiguity of abnormal regions, which reduces isolated false positives and false negatives in the final detection result. Experiments on four real-world medical imaging datasets demonstrate that InvDetect consistently outperforms eighteen prior unsupervised anomaly detection methods.

Keywords: Unsupervised anomaly detection · DDIM inversion

1 Introduction

Anomaly detection in medical imaging plays an important role in many real-world clinical workflows, such as early disease diagnosis, surgical planning, and treatment monitoring [5, 9]. Existing works can be broadly categorized into supervised, self-supervised, and unsupervised methods. Supervised methods [3] rely on pixel-level annotations of abnormal regions, which are costly and time-consuming to obtain. Self-supervised methods [9] attempt to reduce annotation

⋆ Corresponding author: Lingyang Chu (chul9@mcmaster.ca)

costs by generating synthetic anomalies for model training, but their reliability is often compromised by the domain discrepancy between synthetic and real pathological patterns [9]. In contrast, unsupervised methods [28, 5, 15, 2, 26] require only normal images for training. This avoids costly image annotation and the domain discrepancy introduced by synthetic anomalies, while enabling better generalization to unseen abnormal patterns [28, 24].

Our work falls in the category of unsupervised methods, which consists of the following two major lines of existing methods.

Reconstruction-based methods [2, 7, 6, 5, 32, 17, 31, 15, 29, 21, 28] use normal images to train a generative model [2, 5, 7, 28]. When detecting anomalies in an input image, they first use the trained model to generate a pseudo-normal image that is similar to the input but without abnormal area. Then, they detect anomalies by thresholding the pixel-wise difference map between the pseudo-normal image and the input. However, generative models often fail to preserve fine pixel-level details, which leads to pixel-level mismatch in normal regions [21]. This produces unreliable difference maps and degrades detection performance. To mitigate this issue, MatchGen [21] performs test-time optimization to refine the pseudo-normal image, while pDDPM [6] adopts patch-wise generation to improve local fidelity. However, the detection performance of these methods remains constrained by pixel-level mismatch, because generative models are designed to capture the distribution of normal images rather than to accurately reproduce the exact pixel values of a specific input image.

Embedding-based methods [11, 30, 10, 24, 20, 13, 26] characterize the distribution of normal embeddings extracted from normal samples (i.e., images or image patches) by using memory banks [10, 20, 24] or probability density models [13]. Given an input sample, an anomaly is detected by measuring how far its embedding deviates from the normal embeddings. However, since the embedding space is not constructed to explicitly separate normal samples from abnormal ones, the embeddings of abnormal samples can become indistinguishable from the normal embeddings, which limits the detection performance of the embedding-based methods.

Is it possible to construct a latent space in which normal and abnormal image patches are well separated, such that anomalies can be effectively detected in the latent space?

In this paper, we propose **InvDetect**, a novel unsupervised anomaly detection method built on a simple yet principled insight: a diffusion model trained exclusively on normal image patches learns denoising dynamics that map Gaussian noise to the manifold of normal patches. When this model is inverted, normal patches map back close to the Gaussian prior in the noise latent space, whereas abnormal patches deviate due to accumulated inversion errors. This asymmetric behavior induces a structured noise latent space in which normal and abnormal patches become better separated. An overview of InvDetect is shown in Fig. 1. The main contributions are summarized as follows.

First, we formalize the construction of a noise latent space for unsupervised anomaly detection via Denoising Diffusion Implicit Model (DDIM) inversion.

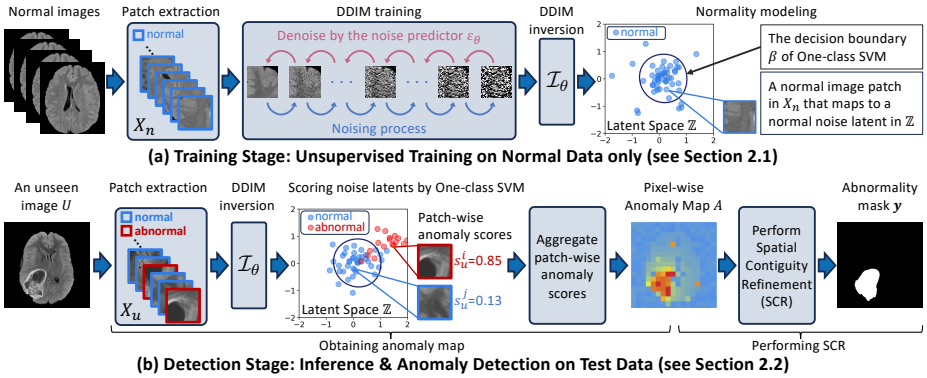

Fig. 1: The training stage and detection stage of InvDetect.

By training the DDIM exclusively on normal image patches, the noise latents of normal patches form a compact cluster in the noise latent space, while those of abnormal patches are more dispersed. This separation property improves the distinguishability between noise latents of normal and abnormal image patches, which enables effective anomaly detection in the noise latent space.

Second, we learn a normality model directly in the noise latent space by training a lightweight one-class SVM on normal noise latents and use it to score unseen image patches. The resulting patch-wise anomaly scores are aggregated into a continuous-valued pixel-wise anomaly map. By performing detection directly in the noise latent space, InvDetect avoids generating pseudo-normal images, which prevents the issue of pixel-level mismatch.

Third, we introduce *spatial contiguity refinement* (SCR), a post-processing module that converts the anomaly map into a binary *abnormality mask* by incorporating a clinically motivated prior that abnormal regions tend to be spatially contiguous. By jointly considering the anomaly map and spatial consistency, SCR suppresses isolated false positives and false negatives, which produces a coherent and high-quality abnormality mask as the final detection result.

Last, we validate InvDetect on four real-world medical imaging datasets, where it achieves strong detection accuracy while maintaining practical detection efficiency compared to eighteen prior unsupervised anomaly detection methods.

2 Our Method: InvDetect

As shown in Fig. 1, InvDetect consists of two stages: a **training stage** that constructs the noise latent space and learns a normality model in this space from normal image patches, and a **detection stage** that produces pixel-wise anomaly scores for unseen images and refines them via spatial contiguity refinement (SCR) to obtain the final binary abnormality mask. We describe the two stages below.

2.1 Training Stage: Noise Latent Space and Normality Modeling

In the training stage (see Fig. 1), InvDetect constructs the noise latent space and learns a normality model from the noise latents of normal image patches. This stage consists of four steps, which are introduced below.

Patch extraction. We extract image patches from each normal training image by a sliding window with size $b \times b$ and a stride of $b/2$ horizontally and vertically. This produces a set of overlapped normal image patches, denoted by $X_n = \{x_n^i\}_{i=1}^M$. The subscript n means "normal". The same patch extraction strategy is applied in the detection stage.

DDIM training. A DDIM defines a diffusion process in which an image is progressively corrupted by Gaussian noise, and a noise predictor ϵ_θ learns to denoise the corruption. In InvDetect, ϵ_θ is trained exclusively on the normal image patches in X_n, so the learned denoising dynamics capture normal image patterns but do not account for abnormal ones. This property causes normal and abnormal patches to behave differently under DDIM inversion, which is essential for inducing a structured noise latent space.

DDIM inversion. Given an image patch $\mathbf{x}$, we apply DDIM inversion to map $\mathbf{x}$ into the noise latent space $\mathbb{Z}$. The inversion follows the deterministic reverse trajectory induced by ϵ_θ to recover the noise latent that would generate $\mathbf{x}$. We denote this mapping as $\hat{\mathbf{z}} = \mathcal{I}_\theta(\mathbf{x})$, where $\hat{\mathbf{z}} \in \mathbb{Z}$ represents the recovered noise latent. Since $\mathcal{I}_\theta$ can be applied to both normal and abnormal image patches, it maps all patches into the noise latent space $\mathbb{Z}$, whose property is analyzed next.

Why does the noise latent space $\mathbb{Z}$ better separate normal and abnormal patches? This separation stems from how ϵ_θ is trained. During DDIM training, noise latents are sampled from a Gaussian prior $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ and mapped to normal image patches. Therefore, when a normal patch $\mathbf{x}_n$ is inverted, its recovered noise latent $\hat{\mathbf{z}}_n = \mathcal{I}_\theta(\mathbf{x}_n)$ remains close to the Gaussian prior, which yields a compact cluster in $\mathbb{Z}$. In contrast, abnormal patches contain patterns that are not captured during training, for which ϵ_θ produces inaccurate noise predictions [28, 12]. When such patches are inverted, mismatch errors accumulate along the deterministic reverse trajectory of ϵ_θ [22], which causes their recovered noise latents $\hat{\mathbf{z}}_a$ to deviate more from the Gaussian prior. Consequently, noise latents inverted from normal and abnormal patches become better separated in $\mathbb{Z}$, which supports effective anomaly detection in this space.

Normality modeling. We apply $\mathcal{I}_\theta$ to all normal image patches in X_n, which yields a set of normal noise latents that forms a compact cluster in $\mathbb{Z}$. Because the latent space is well structured by DDIM training and inversion, normality in $\mathbb{Z}$ can be captured by a simple boundary model. We therefore train a light-weight one-class SVM [25] to learn a *decision boundary* β that encloses the cluster. The one-class SVM is a classical non-parametric estimator of normality and imposes no parametric assumption on the latent distribution. Nonetheless, as detection is performed in the well-structured space $\mathbb{Z}$, InvDetect is not tied to this choice and remains compatible with other one-class estimators, such as parametric Gaussian density modeling [10]. In the detection stage, the anomaly score of an image patch is computed directly in $\mathbb{Z}$ based on the distance of its

noise latent to β, where the patches whose noise latents lie on β are scored zero, and those farther inside/outside β receive smaller/larger anomaly scores.

2.2 Detection Stage: Obtaining Anomaly Map and Performing SCR

In the detection stage (see Fig. 1), InvDetect first produces a pixel-wise anomaly map by aggregating patch-wise anomaly scores computed in the noise latent space $\mathbb{Z}$, and then refines this map via spatial contiguity refinement (SCR) to obtain the final binary abnormality mask.

Obtaining anomaly map. Given an unseen image U, we extract overlapped patches $X_u = \{\mathbf{x}_u^i\}_{i=1}^N$ using the same strategy as in the training stage, where the subscript u means "unseen". Each patch $\mathbf{x}_u^i$ is mapped to its noise latent $\hat{\mathbf{z}}_u^i = \mathcal{I}_\theta(\mathbf{x}_u^i)$, which is assigned a *patch-wise anomaly score* s_u^i by the one-class SVM in $\mathbb{Z}$. These patch-wise anomaly scores $\{s_u^i\}_{i=1}^N$ are aggregated into a *pixel-wise anomaly map* A. Specifically, denote by p a pixel in U and by $D_p \subseteq X_u$ the set of patches containing p. The anomaly score at p is given by

$$A(p) = \sum_{\mathbf{x}_u^i \in D_p} w_p^i s_u^i / \sum_{\mathbf{x}_u^i \in D_p} w_p^i, \tag{1}$$

which is a weighted average of the anomaly scores of all the image patches in D_p. The weight $w_p^i = \exp\big(-\|p - c_u^i\|_2^2/\omega^2\big)$ measures the proximity between p and the center c_u^i of patch $\mathbf{x}_u^i$, and $\omega = b/4$. This projects the anomaly scores obtained in $\mathbb{Z}$ back to the anomaly map in image domain.

Performing spatial contiguity refinement (SCR). Although the anomaly map A provides pixel-wise anomaly scores collected from $\mathbb{Z}$, it does not explicitly model the spatial contiguity of abnormal regions. Direct thresholding may therefore produce fragmented abnormality masks. To incorporate the spatial contiguity prior, we formulate SCR as a global energy minimization problem:

$$\arg\min_{\mathbf{y}} \sum_{p \in U} \Phi\big(y_p; A(p)\big) + \lambda \sum_{(p,q) \in \mathcal{N}} \Psi(y_p, y_q), \tag{2}$$

where $\mathbf{y} = \{y_p \in \{0,1\} \mid p \in U\}$ denotes the *abnormality mask*, and the *abnormality labels* $y_p = 1$ and $y_p = 0$ indicate abnormal and normal pixels, respectively. The **first term** enforces consistency between the binary label y_p and the anomaly score $A(p)$ through a cross-entropy formulation:

$$\Phi\big(y_p; A(p)\big) = -y_p \log \sigma\big(A(p)\big) - (1 - y_p) \log\big(1 - \sigma\big(A(p)\big)\big), \tag{3}$$

where $\sigma(\cdot)$ is the sigmoid function. This term translates the continuous anomaly scores in A into a probabilistic labeling cost, which ensures that the final abnormality mask remains consistent with these anomaly scores. The **second term** introduces a spatial contiguity prior by penalizing label discontinuities between neighboring pixels. Specifically,

$$\Psi(y_p, y_q) = \alpha_{p,q}\, \mathbf{1}(y_p \neq y_q), \tag{4}$$

Table 1: The detailed information of the datasets.

Datasets	Modality	$\mathcal{D}_{train}$	$\mathcal{D}_{test}$	Image Type	Resolution
BraTS2021 [4, 1]	Brain MRI	4,500	500	grayscale	128×128×1
BTCV [19, 1]	Liver CT	3,200	500	grayscale	512×512×1
RESC [14]	Retinal OCT	6,200	500	grayscale	256×256×1
IDRiD [23, 1]	Fundus Images	7,000	500	color	128×128×3

where $(p, q) \in \mathcal{N}$ denotes 4-neighborhood adjacency and $\lambda > 0$ balances the two terms. A larger λ suppresses more isolated false positives but may also cause small abnormal regions to be missed. The **penalization weight** is defined as $\alpha_{p,q} = \exp\left(-\|\mathbf{r}_p - \mathbf{r}_q\|_2^2/\tau^2\right)$, where $\mathbf{r}_p = \sum_{\mathbf{x}_u^i \in D_p} w_p^i \hat{\mathbf{z}}_u^i / \sum_{\mathbf{x}_u^i \in D_p} w_p^i$ is the weighted aggregation of the noise latents whose patches contain pixel p. By defining the penalization weight based on the noise latents in $\mathbb{Z}$, SCR enforces spatial contiguity according to the latent similarity. Neighboring pixels with similar latent representations are therefore encouraged to share the same abnormality label, while structural label transitions are preserved when their latent representations differ.

The energy minimization problem is solved by the Boykov–Kolmogorov max-flow/min-cut algorithm [8, 18], which yields a globally optimal abnormality mask under this formulation.

3 Experiments

In this section, we first introduce the experiment settings and then discuss the experimental results in Sections 3.1 and 3.2. All experiments are conducted on a desktop with RTX 4090 GPU, Intel Core-i9 CPU and 32 GB RAM.

Datasets. We use the public datasets listed in Table 1. The training dataset $\mathcal{D}_{train}$ contains only normal images used for training. The testing dataset $\mathcal{D}_{test}$ contains pixel-wisely annotated abnormal images, which are only used for performance evaluation. We thank the original authors for the excellent datasets.

Evaluation Metrics. Following the literature [5, 21, 9], we use *Dice* to measure overlap between detected and ground-truth abnormal regions, and use *Precision* to measure the fraction of detected abnormal pixels that are correct. For both the metrics, a larger value means a better performance. We also evaluate detection efficiency by the average *runtime* for detection per image.

Baseline Methods. As listed in Table 2, we focus on comparing with the two major category of unsupervised anomaly detection methods, including reconstruction-based methods and embedding-based methods. For all the baselines, we use the official code in default and parameters released by the authors.

Implementation Details of InvDetect. Following [27], the noise predictor ϵ_θ adopts a time-conditioned U-Net with four encoder–decoder stages and is trained using the Adam optimizer [16] with a learning rate of 10^{-3}. The diffusion process uses $T = 1000$ timesteps, and each image patch is mapped to its noise

Table 2: Dice and Precision (Prec.) results. 1st-place in **bold** and 2nd underlined.

	Method	BraTS2021		BTCV		RESC		IDRiD	
		Dice	Prec.	Dice	Prec.	Dice	Prec.	Dice	Prec.
Reconstruction-based	AnoDDPM [28]	51.63	47.67	26.39	19.53	24.07	15.58	19.76	12.43
	THOR [7]	52.71	49.83	21.27	17.75	26.40	16.78	14.64	8.04
	AE-ℓ_1 [5]	34.08	24.67	20.99	16.29	26.04	16.28	21.22	13.69
	CeAE [32]	32.71	23.70	20.18	14.78	26.98	16.93	23.86	19.02
	VAE [17]	38.12	28.27	21.45	16.67	25.57	15.98	26.43	20.27
	VAE-Grad [31]	36.62	29.03	21.11	16.80	26.41	16.62	25.72	19.41
	DAE [15]	59.22	52.83	25.99	19.28	26.14	16.17	15.22	8.69
	GANomaly [2]	35.61	26.57	24.71	18.15	23.68	15.25	15.23	8.93
	DDGAN [29]	25.06	17.73	20.23	15.16	25.84	16.01	22.06	18.13
	MatchGen-GANomaly [21]	40.71	34.43	27.47	21.71	27.21	17.14	17.34	9.52
	MatchGen-DAE [21]	67.16	70.47	28.14	22.85	29.74	18.70	16.68	11.10
Embeding-based	RD4AD [11]	27.24	18.71	10.53	6.82	28.47	19.78	28.95	24.92
	STFPM [30]	30.16	23.47	11.29	7.54	34.51	27.20	26.54	20.77
	PaDiM [10]	26.41	18.80	5.96	3.88	33.21	22.88	29.39	25.71
	PatchCore [24]	32.56	23.75	14.05	8.67	40.25	29.97	28.63	24.75
	CFA [20]	29.69	20.46	15.26	9.52	32.49	22.03	26.12	20.82
	CFLOW [13]	16.12	11.05	4.59	3.42	37.58	30.13	20.49	14.25
	ADINO-DPMM [26]	31.12	22.64	10.13	6.05	36.67	29.78	25.12	20.28
Ours	InvDetect w/o SCR	64.74	65.26	31.55	22.91	40.87	31.14	28.84	24.55
	InvDetect	**68.35**	**70.66**	**32.83**	**24.86**	**42.88**	**34.62**	**33.35**	**31.03**

latent by the DDIM inversion $\mathcal{I}_\theta$ with 200 steps. For the one-class SVM, we use an RBF kernel with $\nu = 0.1$, and set γ to the inverse of the product of the dimension of $\mathbb{Z}$ and the variance of the normal noise latents. We set patch size to $b = 32$, thus the stride is $b/2 = 16$ and $\omega = b/4 = 8$. We set τ to match the mean of $\|\mathbf{r}_p - \mathbf{r}_q\|_2^2$ on $\mathcal{D}_{train}$, and follow [24, 26] to select λ as the smallest value that keeps the detected abnormal area (i.e., false positives) below 1% on the normal images in $\mathcal{D}_{train}$. We uniformly set $\tau = 10$ and $\lambda = 25$ across all datasets. To assess the effect of SCR, we also implement an ablated version of InvDetect w/o SCR, which obtains abnormality mask $\mathbf{y}$ by thresholding A at a threshold $\eta = 0$. Our code is available at https://github.com/lele0007/InvDetect.

3.1 Detection Performance: Dice and Precision

Table 2 compares the Dice and Precision between our methods and the baseline methods. We can draw the following conclusions from the results.

First, for our methods, InvDetect consistently outperforms InvDetect w/o SCR, which shows the effectiveness of SCR in improving detection performance.

Second, the reconstruction-based methods are inferior to InvDetect due to pixel-level mismatch [28]. By performing test-time optimization to reduce pixel-level mismatch, MatchGen-DAE achieves the 2nd-place on BraTS2021, however, it remains inferior to our methods on the other datasets.

Third, embedding-based methods are inferior to InvDetect, because their embedding spaces are not constructed to explicitly separate normal images from abnormal ones. PaDim achieves 2nd-place on IDRiD, but it is consistently outperformed by our methods on the other datasets.

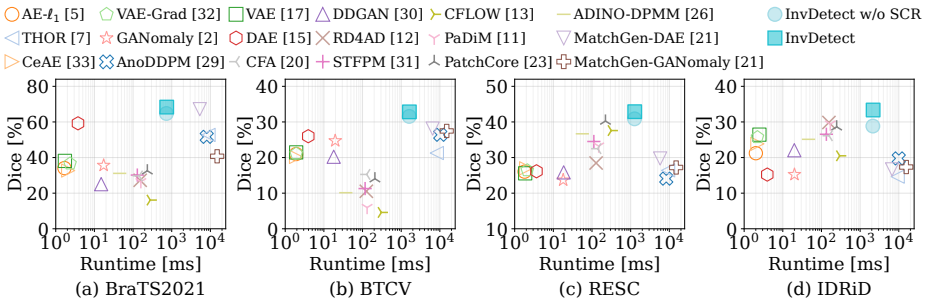

Fig. 2: The results of average runtime (miliseconds) vs. Dice. Reconstruction-based methods are shown by unfilled area-based markers (e.g., □, △, ☆, etc). Embedding-based methods are shown by stroke-based markers (e.g., ×, +, etc).

In summary, InvDetect achieves the best performance across all datasets because: 1) it does not generate pseudo-normal images, which avoids pixel-level mismatch; 2) the noise latent space $\mathbb{Z}$ provides better separation between normal and abnormal image patches, which yields high-quality patch-wise anomaly maps; and 3) SCR incorporates spatial contiguity priors by leveraging the noise latents in $\mathbb{Z}$, which suppresses isolated false positives and false negatives, and produces more coherent abnormality masks.

3.2 Detection Efficiency: Runtime v.s. Dice

Fig. 2 shows the performance of runtime vs. Dice of all the methods. We can draw the following conclusions from the results.

First, the reconstruction-based methods [28, 7, 5, 32, 17, 31, 15, 2, 29, 21] that rely on light-weight image generation models achieve the smallest runtime of $10^0 \sim 10^1$ milliseconds (ms), but their low-quality image generation introduces significant pixel-level mismatch, which limits the Dice score. In contrast, the reconstruction-based methods [28, 7, 5, 32, 17, 31, 15, 2, 29, 21] that employ computationally intensive generation models often achieve higher Dice scores, but at the cost of a much larger runtime around 10^4 ms.

Second, the embedding-based methods [11, 30, 10, 24, 20, 13, 26] achieve small runtime around 10^2 ms, because they work on image embeddings that are efficient to obtain by an encoder network (e.g., ResNet-50 or DINOv2). However, they cannot consistently achieve good Dice scores on all datasets because their embedding spaces are not constructed to separate normal and abnormal images.

Last, InvDetect consistently achieves the highest Dice while maintaining a runtime around two seconds per image, which remains competitive among baselines with high Dice. The runtime of InvDetect w/o SCR is close to InvDetect, because the time cost of SCR is negligible compared to the DDIM inversion $\mathcal{I}_\theta$. Although $\mathcal{I}_\theta$ dominates the computation, InvDetect remains efficient because mapping many image patches by $\mathcal{I}_\theta$ can be parallelized on GPUs, and the total time cost scales linearly with the number of patches.

4 Conclusion

In this paper, we proposed InvDetect, an unsupervised anomaly detection method that operates in a structured noise latent space induced by DDIM inversion. By training the DDIM exclusively on normal image patches, InvDetect induces a noise latent space in which normal and abnormal patches become better separated. Anomaly detection is performed directly in this latent space using a lightweight one-class SVM, which avoids pseudo-normal image generation and the associated pixel-level mismatch. To account for the spatial contiguity of pathological regions, we further incorporate a spatial contiguity prior through a global energy minimization framework, which produces coherent abnormality masks. Extensive experiments on four medical imaging datasets demonstrate that InvDetect achieves strong detection accuracy while maintaining practical detection efficiency.

Acknowledgments. This work is supported in part by the NSERC Discovery Grant program (RGPIN-2022-04977), the Canadian Institutes of Health Research (CIHR) (Funding Reference Number: ACD 187254), and the NSERC sMAP CREATE grant #542989-2020. Jian Pei's research was supported in part by the NSF Project MSPA-2434666. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Creative Commons Attribution 4.0 International (CC BY 4.0) License: https://creativecommons.org/licenses/by/4.0
2. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training. In: Asian Conference on Computer Vision. pp. 622–637 (2019)
3. Asgari Taghanaki, S., Abhishek, K., Cohen, J.P., Cohen-Adad, J., Hamarneh, G.: Deep Semantic Segmentation of Natural and Medical Images: A Review. Artificial Intelligence Review **54**, 137–178 (2021)
4. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv preprint arXiv:2107.02314 (2021)
5. Baur, C., Denner, S., Wiestler, B., Navab, N., Albarqouni, S.: Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study. Medical Image Analysis **69**, 101952 (2021)
6. Behrendt, F., Bhattacharya, D., Krüger, J., Opfer, R., Schlaefer, A.: Patched Diffusion Models for Unsupervised Anomaly Detection in Brain MRI. In: Medical Imaging with Deep Learning. pp. 1019–1032 (2024)
7. Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.: Diffusion Models with Implicit Guidance for Medical Anomaly Detection. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 211–220 (2024)

8. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**, 1124–1137 (2004)
9. Cao, Q., Fan, J., Cai, W.: ART-ASyn: Anatomy-aware Realistic Texture-based Anomaly Synthesis Framework for Chest X-Rays. arXiv preprint arXiv:2512.00310 (2025)
10. Defard, T., Setkov, A., Loesch, A., Audigier, R.: PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization. In: International Conference on Pattern Recognition. pp. 475–489 (2021)
11. Deng, H., Li, X.: Anomaly Detection via Reverse Distillation from One-Class Embedding. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)
12. Graham, M.S., Pinaya, W.H., Tudosiu, P.D., Nachev, P., Ourselin, S., Cardoso, J.: Denoising diffusion models for out-of-distribution detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2948–2957 (2023)
13. Gudovskiy, D., Ishizaka, S., Kozuka, K.: CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows. In: IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 98–107 (2022)
14. Hu, J., Chen, Y., Yi, Z.: Automated Segmentation of Macular Edema in OCT Using Deep Neural Networks. Medical Image Analysis **55**, 216–227 (2019)
15. Kascenas, A., Pugeault, N., O'Neil, A.Q.: Denoising Autoencoders for Unsupervised Anomaly Detection in Brain MRI. In: International Conference on Medical Imaging with Deep Learning. pp. 653–664 (2022)
16. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 (2014)
17. Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes. In: International Conference on Learning Representations (2014)
18. Kolmogorov, V., Zabin, R.: What Energy Functions Can Be Minimized via Graph Cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence **26**, 147–159 (2004)
19. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: MICCAI Multi-Atlas Labeling Beyond the Cranial Vault–Workshop and Challenge. In: International Conference on Medical Image Computing and Computer Assisted Intervention Workshop. p. 12 (2015)
20. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access **10**, 78446–78454 (2022)
21. Ma, X., Ma, J., He, S., Che, X., So, H.Y., Chu, L.: MatchGen: Detecting Medical Abnormal Region by Generating Matched Normal Regions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 321–331 (2025)
22. Mokady, R., Hertz, A., Aberman, K., Pritch, Y., Cohen-Or, D.: Null-text inversion for editing real images using guided diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6038–6047 (2023)
23. Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., et al.: IDRiD: Diabetic Retinopathy–Segmentation and Grading Challenge. Medical Image Analysis **59**, 101561 (2020)

24. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards Total Recall in Industrial Anomaly Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2022)
25. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the Support of a High-Dimensional Distribution. Neural Computation **13**, 1443–1471 (2001)
26. Schulthess, N., Konukoglu, E.: Anomaly Detection by Clustering DINO Embeddings Using a Dirichlet Process Mixture. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 46–56 (2025)
27. Song, J., Meng, C., Ermon, S.: Denoising Diffusion Implicit Models. arXiv preprint arXiv:2010.02502 (2020)
28. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 650–656 (2022)
29. Xiao, Z., Kreis, K., Vahdat, A.: Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. arXiv preprint arXiv:2112.07804 (2021)
30. Yamada, S., Hotta, K.: Reconstruction Student with Attention for Student-Teacher Pyramid Matching. arXiv preprint arXiv:2111.15376 (2021)
31. Zimmerer, D., Isensee, F., Petersen, J., Kohl, S., Maier-Hein, K.: Unsupervised Anomaly Localization Using Variational Auto-Encoders. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 289–297 (2019)
32. Zimmerer, D., Kohl, S.A., Petersen, J., Isensee, F., Maier-Hein, K.H.: Context-encoding Variational Autoencoder for Unsupervised Anomaly Detection. arXiv preprint arXiv:1812.05941 (2018)