

# Diffusion Models with Implicit Guidance for Medical Anomaly Detection

Cosmin I. Bercea<sup>1,2</sup>, Benedikt Wiestler<sup>3</sup>, Daniel Rueckert<sup>1,3,5</sup>, and Julia A. Schnabel<sup>1,2,4</sup>

<sup>1</sup> Technical University of Munich, Munich, Germany

<sup>2</sup> Helmholtz AI and Helmholtz Center Munch, Munich, Germany

<sup>3</sup> Klinikum Rechts der Isar, Munich, Germany

<sup>4</sup> Kings College London, London, UK

<sup>5</sup> Imperial College London, London, UK

**Abstract.** Diffusion models have advanced unsupervised anomaly detection by improving the transformation of pathological images into pseudo-healthy equivalents. Nonetheless, standard approaches may compromise critical information during pathology removal, leading to restorations that do not align with unaffected regions in the original scans. Such discrepancies can inadvertently increase false positive rates and reduce specificity, complicating radiological evaluations. This paper introduces Temporal Harmonization for Optimal Restoration (*THOR*), which refines the de-noising process by integrating implicit guidance through temporal anomaly maps. *THOR* aims to preserve the integrity of healthy tissue in areas unaffected by pathology. Comparative evaluations show that *THOR* surpasses existing diffusion-based methods in detecting and segmenting anomalies in brain MRIs and wrist X-rays. Code: [https://github.com/ci-ber/THOR\\_DDPM](https://github.com/ci-ber/THOR_DDPM).

**Keywords:** Generative AI · Representation Learning · Normative Learning · OoD Detection · Medical Image Analysis · Machine Learning

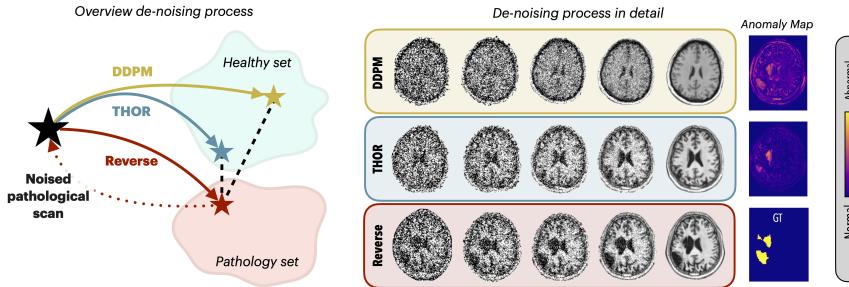


Fig. 1: Denoising diffusion probabilistic models (DDPMs) trend towards a generalized healthy reference. From a noised starting point, *THOR* employs temporal harmonization to yield an outcome resembling the original in healthy tissue regions, concurrently replacing pathology areas with pseudo-healthy restorations.

## 1 Introduction

Robust and accurate anomaly detection is vital for early diagnosis and effective treatment, especially in the face of rare and diverse pathologies. The complexity and variability inherent in medical conditions present substantial challenges to conventional diagnostic methods anchored in supervised learning [1,2]. These methods depend heavily on extensive, annotated datasets, which are difficult to obtain for rare conditions, limiting the scope and flexibility of diagnostic tools. In response, unsupervised learning has emerged as a viable alternative, capable of detecting anomalies across a broad spectrum without the need for explicit labels [3,4,5,6]. Among unsupervised techniques, denoising diffusion probabilistic models (DDPMs) [7] have shown substantial promise in enhancing the precision and efficiency of anomaly detection. By adding and subsequently removing noise, DDPMs transform pathological inputs into pseudo-healthy outputs, demonstrating impressive generative potential. Nonetheless, this noise-dependent process can result in significant loss of information, leading restored images to deviate from their original state, including in regions unaffected by pathology. Such deviations risk increasing false positives and decreasing specificity, further complicating the diagnostic process.

To overcome the limitations inherent in DDPMs, more sophisticated models have been developed. AnoDDPM proposes to use Simplex noise, which allows the use of lower noise levels [8]. Conditional diffusion models blend the capabilities of autoencoders with diffusion techniques to incorporating semantic information such as tissue intensity into the de-noising process [9]. Patch-based DDPMs (pDDPMs) extend these advancements by applying the diffusion process to localized patches of the image, using adjacent areas as contextual anchors in a sliding-window technique [10]. AutoDDPMs build upon this foundation with a unique approach that involves masking, stitching, and re-sampling, utilizing dual de-noising processes at different levels of noise to seamlessly integrate context into the reconstructions [11]. While these innovations represent substantial progress, they also introduce complexities. The task of determining an optimal patch size that can adapt to the multiple scales of diseases is challenging due to the diversity of pathological presentations. Additionally, the complexity of orchestrating dual de-noising processes across different noise levels requires precise calibration. These challenges could potentially limit their practicability.

Diffusion models enhanced with classifier guidance use weakly supervised classifiers for anomaly detection, leveraging gradients to refine the identification of anomalous regions [12]. However, the effectiveness of this approach depends on the accuracy of classifiers, potentially limiting its capability to detect diseases independently by biasing it towards known pathologies.

In this work, we introduce *THOR* (Temporal Harmonization for Optimal Restoration), a novel approach designed to enhance unsupervised anomaly detection in medical imaging, as illustrated in Fig. 1. *THOR* incorporates implicit guidance into diffusion models through the use of temporal anomaly maps, aiming to preserve the original image context while achieving accurate anomaly detection and segmentation. Our key contributions are as follows:

- We develop *THOR* and leverage implicit guidance within diffusion models to facilitate optimal image restorations and improve the accuracy of anomaly segmentation.
- We apply *THOR* to two challenging medical datasets, where it demonstrates its capability in accurately segmenting stroke lesions and localizing pathology in pediatric wrist X-rays, thereby enhancing performance in essential diagnostic tasks.
- We perform a sensitivity analysis of critical hyperparameters such as different noise types and levels.

## 2 Background

### 2.1 Anomaly Detection Setup

In medical imaging anomaly detection, the objective is to detect deviations from normal anatomical structures without explicit pathological labels. We define  $\tilde{X}$  as the domain of all medical images, where each image  $\tilde{x} \in \tilde{X}$  includes regions of both normal and abnormal tissue. The aim is to assign an anomaly score  $S$  to each pixel (or voxel), using a function  $f : \tilde{X} \rightarrow S$ .

Considering a dataset of medical images  $x_{i=1}^N \in X \subset \tilde{X}$  for training, these images are presumed to represent a healthy tissue distribution, denoted as  $P(X)$ . The challenge lies in accurately modeling  $P(X)$ . By doing so, we can project any input image into the  $P(X)$  space, creating a pseudo-healthy reconstruction. If an input image has pathology, this method produces a version where pathological features are replaced with those typical of healthy tissue according to  $P(X)$ . This approach enables anomaly detection by contrasting the original image with its pseudo-healthy counterpart to identify deviations.

### 2.2 Denoising Diffusion Probabilistic Models (DDPMs)

DDPMs are generative models that aim to replicate the distribution  $P(X)$  through a process that incrementally introduces and reverses Gaussian noise.

**Forward Process.** In the forward process, a DDPM gradually transforms a clean image  $x_0$ , drawn from the distribution  $P(X)$ , into a completely noisy state over a Markov chain of  $T$  steps, described by:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad (1)$$

where  $\alpha_t$  is part of a predetermined noise schedule ( $0 < \alpha_t < 1$ ), and  $\epsilon$  represents Gaussian noise. The sequence  $x_{t=0}^T$  depicts the transition of the input image into a state where  $x_T$  is predominantly noise.

**Reverse Process.** The reverse process aims to reconstruct the clean image from its noisy counterpart by denoising, essentially learning  $P(X)$ . This can be formulated as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} (x_t - \sqrt{1 - \alpha_t} \epsilon_\theta(x_t, t)), \quad (2)$$

where  $\epsilon_\theta(x_t, t)$  is the estimate of the noise added at step  $t$ . By learning  $\epsilon_\theta$ , the model inverts the noising process, approximating the distribution of  $P(X)$ .

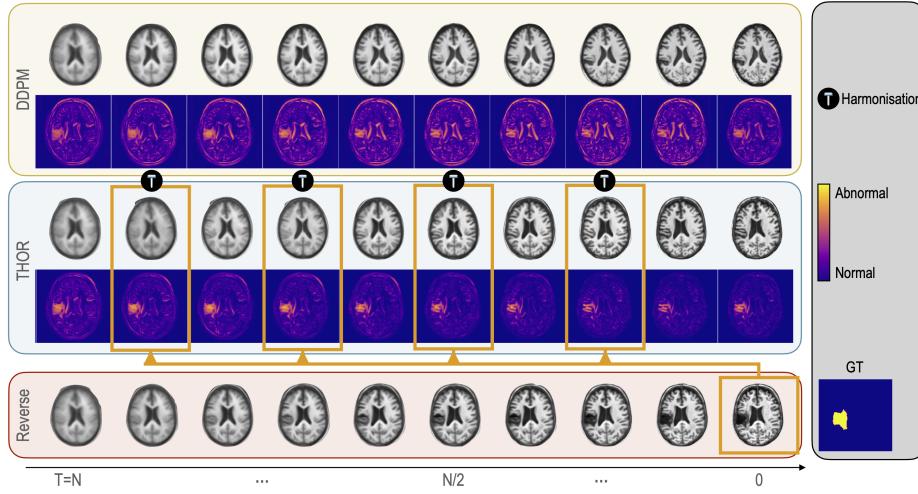


Fig. 2: The top row displays the traditional DDPM de-noising sequence, where noise is progressively reduced to clarify image features. In contrast, the middle row showcases *THOR*: starting with equivalent high noise levels and then strategically applying unsupervised temporal anomaly masks at key intervals (indicated by orange borders) to 'harmonize' the image. This 'harmonization' process selectively refines the image by maintaining normal tissue integrity while attenuating anomalies. The bottom row shows the reverse process with anomalies becoming increasingly apparent as noise is reversed, culminating in the ground truth (GT) image where the anomaly is clearly delineated.

### 3 Method: THOR

*THOR* advances the de-noising process in DDPMs, offering guidance during inference through the application of unsupervised temporal anomaly masks, without necessitating retraining. Typically, DDPMs necessitate high noise levels ( $T$ ) to effectively obscure anomalies, a practice that can compromise the integrity of non-pathological tissue details. Such an approach may result in the loss of critical anatomical information, thereby elevating the potential for false positives. The innovation of *THOR* lies in its ability to guide the restoration process by strategically reintegrating healthy tissue information, a technique we refer to as "harmonization." This method starts at the same elevated noise levels but diverges by using implicit temporal anomaly masks to inform the denoising trajectory. Such guidance aims to selectively restore the image, focusing on preserving the fidelity of non-pathological regions while reducing the anomalies. Details of our procedural approach are delineated in Fig. 2.

**Implicit Guidance via Intermediate Anomaly Maps.** Intermediate anomaly maps play an essential role in the unsupervised "harmonization" process of *THOR*, applied at specific timesteps. These maps critically compare the predic-

tive reconstructions  $x_0^t$  with the actual input image  $x_0^{\text{input}}$ , highlighting discrepancies that indicate anomalies and distinguishing regions that are likely healthy. Anomaly maps  $m$  combine residual differences with the Learned Perceptual Image Patch Similarity (LPIPS) metric [13], enhancing the identification of subtle pathological changes [3]:

$$m(x, x_{rec}) = |x - x_{rec}| \cdot S_{\text{LPIPS}}(x, x_{rec}). \quad (3)$$

To avoid incorporating anomalous regions in the denoising process, we normalize the values of  $m$  between 0 and 1 and apply morphological operations, specifically a sequence of closing followed by dilation (denoted as  $cd$ ):

These anomaly maps are then utilized in the "harmonization" process to adjust the interpolation between the pseudo-healthy predictions and the actual inputs. This adjustment aims to produce reconstructions that not only closely resemble the original images but also conform to the healthy tissue profile:

$$x_t = cd(m(x_0^t, x_0^{\text{input}})) \cdot x_0^{\text{prediction}} + (1 - cd(m(x_0^t, x_0^{\text{input}}))) \cdot x_0^{\text{input}}. \quad (4)$$

The final anomaly score,  $S$ , is calculated using the harmonic mean of the anomaly maps at the selected timesteps, explicitly defined as:

$$S = n \left/ \sum_{t \in \text{selected steps}} \frac{1}{m(x_0^t, x_0^{\text{input}})} \right., \quad (5)$$

where  $n$  is the total number of selected harmonization timesteps.

## 4 Experiments and Results

To demonstrate the utility, generalizability, and performance of our method, we conduct two experiments, i.e., the segmentation of ischemic stroke lesion in MRI in Sec. 4.1 and anomaly localization in pediatric wrist X-ray images in Sec. 4.2.

### 4.1 Ischemic Stroke Lesion Segmentation in Brain MRI

Stroke represents a major cause of disability and mortality worldwide, with its early detection being paramount for effective treatment planning.

**Datasets.** The training dataset encompasses 582 T1-weighted MRI scans from the IXI [14] dataset and 217 healthy samples from the ATLAS v2.0 dataset [15], offering a wide representation of normal brain anatomy. For testing, we employed the ATLAS dataset, which includes 655 T1w MRI scans with expertly segmented lesion masks. We categorized the anomalies into small (less than 71 pixels), medium, and large ( $\geq 570$ ) lesions for detailed analysis, excluding 20 slices with significant unannotated hypo-intense artifacts to maintain data integrity. Scans were normalized to the 98th percentile and resized to  $128 \times 128$  pixels, with lesion segmentation evaluated via the maximum achievable Dice.

Table 1: **Performance on Brain MRI Stroke Segmentation.** *THOR*, our proposed method, considerably outperforms other methods (DDPM, AutoDDPM, AnoDDPM, pDDPM) across different lesion sizes, marked by the ***bold*** numbers and percentage improvements ( $\Delta \pm$ ) compared to the best baseline.

Noise	Method	Pathology [Dice] $\uparrow$			Large
		Average	Small	Medium	
Gauss	THOR (ours)	<b>20.41</b> $\Delta 20\%$	<b>9.14</b> $\Delta 103\%$	<b>26.34</b> $\Delta 19\%$	41.26 $\nabla 5\%$
	DDPM [7]	8.05 $\nabla 61\%$	1.37 $\nabla 85\%$	9.53 $\nabla 64\%$	25.65 $\nabla 38\%$
	AutoDDPM [11]	16.95 $\nabla 17\%$	4.55 $\nabla 50\%$	22.07 $\nabla 16\%$	<b>43.47</b> $\Delta 5\%$
Simplex	THOR (ours)	<b>29.74</b> $\Delta 33\%$	<b>11.54</b> $\Delta 44\%$	<b>39.20</b> $\Delta 30\%$	<b>63.64</b> $\Delta 34\%$
	AnoDDPM [8]	18.07 $\nabla 39\%$	4.82 $\nabla 58\%$	23.45 $\nabla 40\%$	46.65 $\nabla 27\%$
	pDDPM [10]	22.28 $\nabla 25\%$	8.02 $\nabla 31\%$	30.16 $\nabla 23\%$	47.66 $\nabla 25\%$

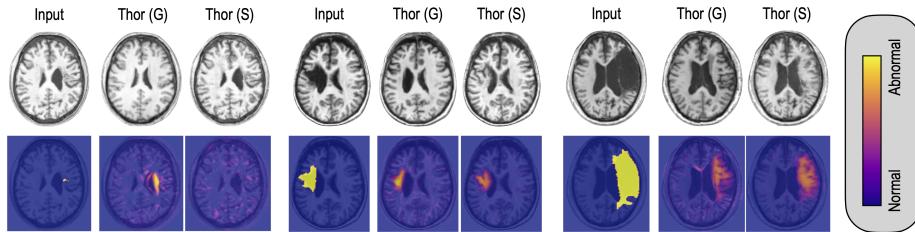


Fig. 3: Anomaly detection in brain MRI scans processed by *THOR* using Gaussian (G) and Simplex (S) noise. From left to right, the lesions increase in size, with the smallest representing a challenging case.

**Results.** Tab. 1 shows quantitative results and explores two key diffusion noise scenarios: Gaussian and Simplex. This examination is vital for assessing the performance of *THOR* in comparison with leading diffusion models. *THOR* is proficient with both types of noise, illustrating its broad applicability.

*Gaussian noise* is the conventional choice for DDPMs but introduces challenges in anomaly detection. Due to the partial denoising strategy employed for anomaly detection, a high noise level (here  $T=350$ ) is essential to effectively conceal anomalies [8]. Yet, deploying Gaussian noise at such high iterations frequently results in false positives due to inaccuracies in restoring healthy tissue. This limitation is reflected in the diminished segmentation scores for DDPM. Conversely, our harmonization process navigates the de-noising towards more precise restorations. Consequently, *THOR* addresses the challenge of false positives and significantly refines the accuracy of anomaly segmentation, as evidenced both numerically in Tab. 1 and visually in Fig. 2 and Fig. 3.

*Simplex noise* provides a notable advantage with its coarse noise patterns, allowing for the de-noising process to commence at lower levels ( $T=250$ ) as demonstrated in [8]. This characteristic is beneficial, preserving more of the original image context and laying a stronger groundwork for restoration. The utility of Simplex noise becomes apparent when observing the improved performance of

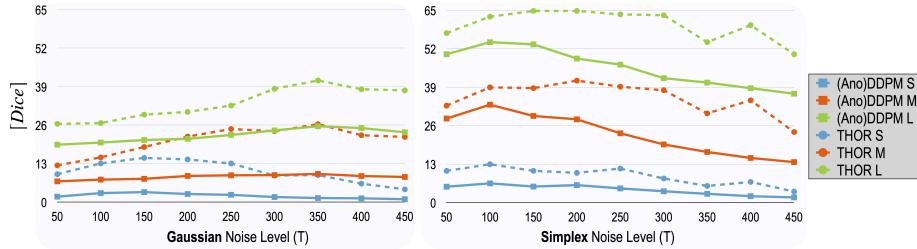


Fig. 4: **Noise Level Ablation.** *THOR* outperforms the diffusion counterparts under both Gaussian and Simplex noise types across different noise levels  $T$ .

models like AnoDDPM, which exhibit significant enhancements over the traditional DDPM. Leveraging the capability of Simplex noise, *THOR* advances the restoration process further. Its harmonization process meticulously refines the output, ensuring restorations more faithfully represent the original healthy tissue and thereby outperforming AnoDDPM and similar models (see Tab. 1).

**Sensitivity analysis of noise levels  $T$**  is shown in Fig. 4. Increasing noise levels in the Gaussian setting enhances the detection of larger lesions, highlighting the role of higher noise in their effective masking. In contrast, performance declines with elevated Simplex noise levels, likely due to a self-supervision effect where anomalies mimicking the coarse noise pattern are identified and eliminated. While it achieves improved Dice scores for stroke segmentation, caution is advised. Its specificity to the coarse noise pattern may limit its effectiveness in broader anomaly detection scenarios by potentially overlooking anomalies that do not match this pattern. *THOR* excels across different noise intensities, showcasing particular robustness at higher levels. This robustness minimizes the need for finely tuned noise adjustments for specific applications or anomaly sizes, underscoring *THOR*'s adaptability and efficacy in anomaly detection tasks.

#### 4.2 Anomaly Localization in Pediatric Wrist X-rays

Bone fractures are notably prevalent in children, with their detection being a critical step in ensuring timely medical intervention.

**Dataset.** We utilize the comprehensive GRAZPEDWRI-DX dataset [16], encompassing 10,643 X-rays of pediatric wrist injuries from 6,091 individual patients. It includes a wide array of anomalies annotated with bounding boxes by certified pediatric radiologists. This includes bone anomalies (BA), foreign bodies (FB), fractures (Frac.), the presence of metal implants, periosteal reactions (PR), and soft tissue conditions (Soft). We report the recall and F1 scores.

**Results.** Tab. 2 and Fig. 5 present both quantitative and qualitative outcomes. For this experiment, we concentrated on Gaussian noise, recognizing from prior

Table 2: Anomaly detection and localization results in pediatric wrist X-rays.

Noise Method	BA		FB		Frac.		Metal		Pr.		Soft	
	Recall	F1										
Gauss	<b>83.33</b>	23.76	<b>75.00</b>	25.00	<b>75.39</b>	<b>16.46</b>	<b>99.76</b>	<b>73.76</b>	<b>76.42</b>	16.64	26.32	10.77
DDPM [7]	32.22	6.35	<b>75.00</b>	29.83	28.53	5.10	86.47	39.66	52.25	9.79	23.68	8.89
AutoDDPM [11]	63.89	<b>23.93</b>	<b>75.00</b>	<b>58.33</b>	45.56	15.84	95.89	72.05	62.29	<b>29.00</b>	<b>31.58</b>	<b>16.45</b>



Fig. 5: Anomaly detection in pediatric wrist X-rays processed by *THOR* using Gaussian noise. False positives arise from unannotated non-pathological changes like unnatural bone positions following fractures or the presence of casts.

analysis that Simplex noise lacks versatility for widespread anomaly detection applications. It tends to replicate anomalies dissimilar to its coarse patterns, such as bone anomalies, foreign bodies, and metal implants, as detailed in the supplementary materials. *THOR* outperforms SOTA diffusion models, considerably improving the number of anomalies detected by up to 65% in case of fractures and achieving almost perfect recall in detecting metal implants.

## 5 Discussions and Conclusion

This paper introduces *THOR*, a diffusion-based framework for anomaly detection in medical imaging, which incorporates a novel harmonization process to enhance the denoising and restoration, thereby improving segmentation accuracy. We rigorously tested *THOR* in two challenging scenarios: detecting strokes in brain MRIs and identifying pediatric wrist injuries in X-rays. Our results show that *THOR* considerably outperforms existing diffusion methods.

However, unsupervised anomaly detection still faces challenges such of false positives due to unannotated non-pathological changes shown in Fig. 3 and Fig. 5. These are correctly identified as anomalies, but not annotated by the radiologists. Furthermore, some conditions like soft tissue anomalies are subtle and difficult to spot on small resolutions. Additionally, we discovered that Simplex noise exhibits a self-supervision effect, introducing a bias in the anticipated anomaly distribution, warranting cautious use in wide-ranging anomaly detection tasks.

Our future efforts will focus on overcoming these obstacles to improve the diagnostic precision and broaden the utility of unsupervised anomaly detection across a diverse array of anomalies, organs, and imaging modalities.

## References

1. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: DeepMedic for brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention BrainLes Workshop. pp. 138–149 (2016)
2. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al.: A foundation model for generalizable disease detection from retinal images. *Nature* pp. 1–8 (2023)
3. Bercea, C.I., Wiestler, B., Rueckert, D., A, S.J.: Generalizing unsupervised anomaly detection: Towards unbiased pathology screening. International Conference on Medical Imaging with Deep Learning (2023)
4. Zimmerer, D., Isensee, F., Petersen, J., Kohl, S., Maier-Hein, K.: Unsupervised anomaly localization using variational auto-encoders. In: Medical Image Computing and Computer Assisted Intervention. pp. 289–297. Springer (2019)
5. Pinaya, W.H., Tudosi, P.D., Gray, R., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J.: Unsupervised brain imaging 3d anomaly detection and segmentation with transformers. *Medical Image Analysis* **79**, 102475 (2022)
6. Tan, J., Hou, B., Day, T., Simpson, J., Rueckert, D., Kainz, B.: Detecting outliers with poisson image interpolation. In: Medical Image Computing and Computer Assisted Intervention. pp. 581–591. Springer (2021)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
8. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anodddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops pp. 650–656 (June 2022)
9. Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., Schlaefer, A.: Guided reconstruction with conditioned diffusion models for unsupervised anomaly detection in brain mrис. *arXiv preprint arXiv:2312.04215* (2023)
10. Behrendt, F., Bhattacharya, D., Krüger, J., Opfer, R., Schlaefer, A.: Patched diffusion models for unsupervised anomaly detection in brain mri. International Conference on Medical Imaging with Deep Learning (2023)
11. Bercea, C.I., Neumayr, M., Rueckert, D., Schnabel, J.A.: Mask, stitch, and resample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models. *arXiv preprint arXiv:2305.19643* (2023)
12. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. *Medical Image Computing and Computer Assisted Intervention* pp. 35–45 (2022)
13. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
14. Ixi dataset. <https://brain-development.org/ixi-dataset/>, accessed: 2023-02-15
15. Liew, S.L., Lo, B.P., Miarnda R. Donnelly, e.a.: A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Scientific Data* **9** (2022)
16. Nagy, E., Janisch, M., Hržić, F., et al.: A pediatric wrist trauma x-ray dataset (grazpedwri-dx) for machine learning. *Scientific Data* **9**, 222 (2022). <https://doi.org/10.1038/s41597-022-01328-z>

## 6 Appendix

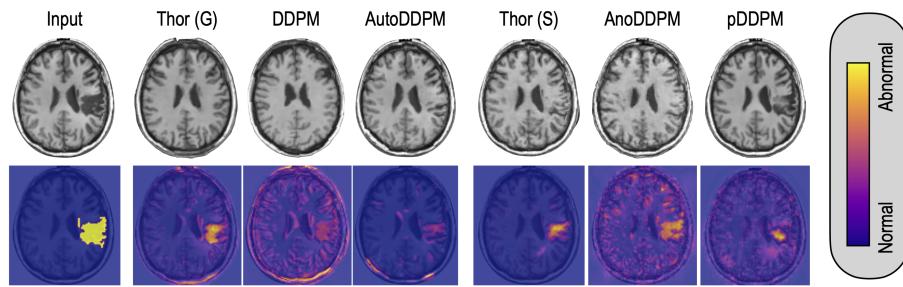


Fig. 6: Qualitative assessment of different diffusion-based models in Brain MRI. *THOR* refines the performance of both DDPM (Gaussian) and AnoDDPM (Simplex), resulting in more accurate reconstructions and enhanced segmentations.

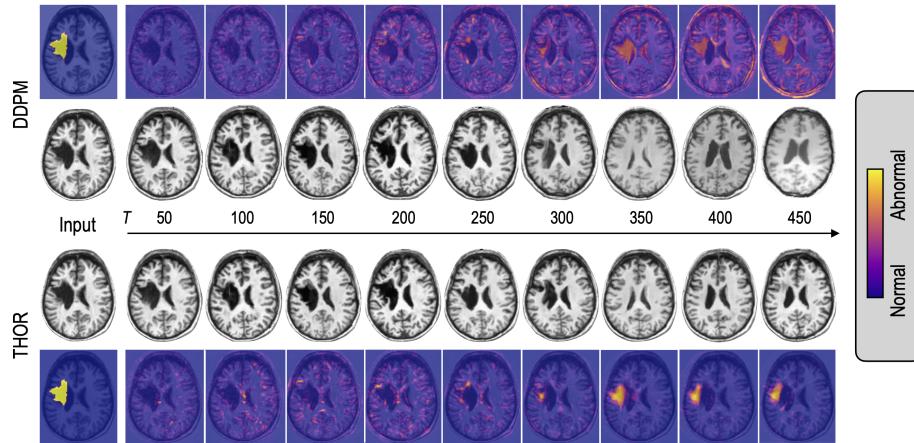


Fig. 7: Comparison of noise scales  $T$  for DDPM and *THOR* using Gaussian noise reveals that low noise levels ( $\leq 300$ ) retain the anomaly's structure, leading to missed detections. At noise levels  $> 300$ , DDPM diverges significantly from the original image, affecting even pathology-free areas. Conversely, *THOR* applies temporal harmonization to generate outputs more closely aligned with the input, yet within the healthy spectrum, across all noise levels  $> 300$ .

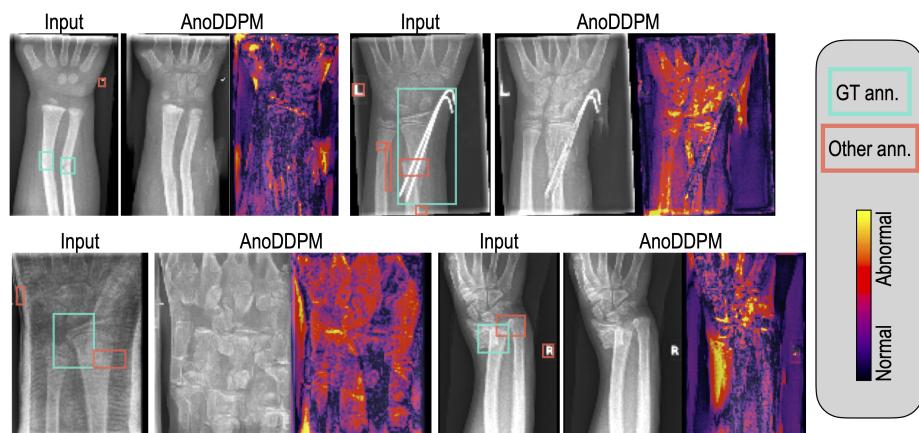


Fig. 8: AnoDDPM’s performance with Simplex noise in detecting anomalies in wrist X-ray images demonstrates that the diffusion process reconstructs fractures, unnatural bone positions, and metal implants. Similar to observations in brain MRI experiments, Simplex noise struggles to accurately learn the healthy anatomy, instead, it aims at eliminating structures akin to its coarse noise patterns. Although literature cites its successful application in tumor segmentation, and our work confirms enhanced performance in stroke lesion segmentation, Simplex noise proves unsuitable for general anomaly detection tasks.