

AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise

Julian Wyatt¹ Adam Leach¹ Sebastian M. Schmon² Chris G. Willcocks¹

Department of {¹Computer Science | ²Mathematical Sciences}, Durham University, Durham, UK

{julian.wyatt, adam.leach, sebastian.schmon, christopher.g.willcocks}@durham.ac.uk

Abstract

Generative models have been shown to provide a powerful mechanism for anomaly detection by learning to model healthy or normal reference data which can subsequently be used as a baseline for scoring anomalies. In this work we consider denoising diffusion probabilistic models (DDPMs) for unsupervised anomaly detection. DDPMs have superior mode coverage over generative adversarial networks (GANs) and higher sample quality than variational autoencoders (VAEs). However, this comes at the expense of poor scalability and increased sampling times due to the long Markov chain sequences required. We observe that within reconstruction-based anomaly detection a full-length Markov chain diffusion is not required. This leads us to develop a novel partial diffusion anomaly detection strategy that scales to high-resolution imagery, named AnoDDPM. A secondary problem is that Gaussian diffusion fails to capture larger anomalies; therefore we develop a multi-scale simplex noise diffusion process that gives control over the target anomaly size. AnoDDPM with simplex noise is shown to significantly outperform both f-AnoGAN and Gaussian diffusion for the tumorous dataset of 22 T1-weighted MRI scans (CCBS Edinburgh) qualitatively and quantitatively (improvement of +25.5% Sørensen–Dice coefficient, +17.6% IoU and +7.4% AUC).

1. Introduction

Denoising diffusion probabilistic models (DDPMs) [7] are capable of generating samples from complex data distributions, with superior mode coverage [27] over generative adversarial networks (GANs) and variational autoencoders (VAEs) [4] (a desirable property for anomaly detection and small datasets). The denoising process takes samples from a $\mathcal{N}(0, I)$ distribution and stochastically transforms them onto a learned data distribution. We leverage this ability and construct a model, AnoDDPM, trained on purely healthy patient data that maps potentially anomalous

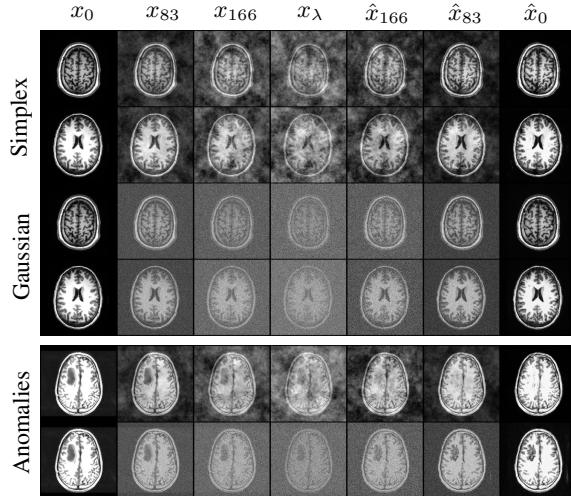


Figure 1. The forward-backward diffusion process where $\lambda = 250$ for simplex noise (rows 1-2) and Gaussian noise (rows 3-4) trained on healthy data. Unlike simplex noise (row 5), Gaussian diffusion fails to repair (row 6) unhealthy anomalous test data.

lous query data onto the healthy distribution through a partial diffusion process. We can then highlight anomalies by comparing with the original image. Partial diffusion allows for better reconstruction of samples from the data distribution and scales to high resolution images well. The shorter Markov chain compared to a full diffusion results in faster inference and training.

Several methods for unsupervised anomaly detection using generative models such as GANs [6] and VAEs [15, 24] have been previously explored. Typically, models learn to generate non-anomalous (healthy) images or regions of interest from anomalous data. Abnormalities are then detected by determining if they lie outside the manifold of the learned representation, or by comparing the generated and original images in pixel space. AnoGAN [21], one of the first GANs used for anomaly detection, scores images based on both the differences in pixel-space and discriminator fea-

tures between a test image and an iteratively refined generator approximation. The refinement process requires several hundred backpropagation iterations, resulting in slow inference. Variations to AnoGAN have been proposed, such as f-AnoGAN [20] and P-AnoGAN [3]. f-AnoGAN utilises a Wasserstein GAN [1] and an encoder to map the query into the latent space to speed up the model. Whereas P-AnoGAN uses a progressive GAN [8] to improve the resolution at which anomalies can be detected. Other approaches such as Zimmerer *et al.* [29] utilise context encoders to inpaint reconstructions of the image and Nguyen *et al.* [10] use super-pixel networks to detect anomalies after the reconstruction as training and image generation can become unstable on larger samples.

This paper proposes a novel anomaly detection approach which utilises DDPMs to corrupt the image and reconstruct a healthy approximation. This provides several advantages over adversarial training as it can better capture smaller datasets with improved sample quality and stable training. However, this approach also comes with unique challenges. As we will demonstrate, simple Gaussian corruption can be inadequate for anomaly detection. As a remedy, we investigate the utility of simplex noise [13]—a popular method in computer graphics—instead of Gaussian noise.

In summary, the main contributions of this paper are:

- A partial diffusion strategy which noises the anomalous image to some parameterised timestep λ , which is reconstructed from the corruption, as seen in Fig. 1.
- Utilising multi-scale (multi-octave) simplex noise to allow larger anomalous regions to become reconstructed as healthy regions.

2. Denoising Diffusion Probabilistic Models

DDPMs [5] have become the state-of-the-art approach for generative modelling with respect to sample quality and mode coverage [27]. DDPMs consist of a forward diffusion process $q(\mathbf{x}_t | \mathbf{x}_{t-1})$ gradually corrupting data from some target distribution $q(\mathbf{x}_0)$ into a normal distribution, and a learned reverse process $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ that generates samples by turning noise into samples from $q(\mathbf{x}_0)$.

The forward process is a non-homogenous Markov chain meaning that the dynamics of the process can be described by the one-step transition density,

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

for $t = 1, \dots, T$. The quantity of noise added at each step is defined by a variance schedule $\beta_t \in (0, 1)$, $t = 1, \dots, T$. This can be defined as a small linear schedule [23], and was updated by Ho *et al.* [7] to increase linearly from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$. Alternatively Dhariwal and Nichol [11]

found a cosine schedule to also perform well. The generative model, parameterised by θ , is the learned reverse process starting with $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$ and sampling according to

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1} | \mu_\theta(\mathbf{x}_t, t), \tilde{\beta}_t \mathbf{I}) \quad (2)$$

for $t = T, \dots, 1$, and $\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t$. μ_θ can, for example, be implemented with U-Net like architectures [17].

The loss function to train $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ can then be taken to be the variational lower bound \mathcal{L}_{vlb} on the marginal likelihood $p_\theta(\mathbf{x}_0)$

$$\mathcal{L}_{vlb} = \mathcal{L}_0 + \mathcal{L}_1 + \dots + \mathcal{L}_{T-1} + \mathcal{L}_T, \quad (3)$$

$$\mathcal{L}_0 = -\log p_\theta(\mathbf{x}_0 | \mathbf{x}_1), \quad (4)$$

$$\mathcal{L}_{t-1} = D_{KL}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) || p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)), \quad (5)$$

$$\mathcal{L}_T = D_{KL}(q(\mathbf{x}_T | \mathbf{x}_0) || p(\mathbf{x}_T)), \quad (6)$$

where D_{KL} denotes the Kullback-Leibler divergence. The above loss is indeed tractable, since $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0)$ has a closed form expression.

In this work, we will consider a modification due to Ho *et al.* [7]: writing $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{i=0}^T \alpha_i$ we have

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t | \mathbf{x}_0 \sqrt{\bar{\alpha}_t}, (1 - \bar{\alpha}_t) \mathbf{I}), \quad (7)$$

$$\mathbf{x}_t = \mathbf{x}_0 \sqrt{\bar{\alpha}_t} + \epsilon_t \sqrt{1 - \bar{\alpha}_t}, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{I}). \quad (8)$$

This allows fast sampling of \mathbf{x}_t for arbitrary t , without having to find intermediate steps $\mathbf{x}_{t-1}, \dots, \mathbf{x}_1$.

Setting

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\alpha_t} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) \quad (9)$$

we can learn ϵ_θ , for example, using a neural network as before. It turns out that Ho *et al.* [7] discovered that the following simplified objective results in better sample quality over \mathcal{L}_{vlb} :

$$\mathcal{L}_s = \mathbb{E}_{t \sim [1-T], \mathbf{x}_0 \sim q(\mathbf{x}_0), \epsilon \sim \mathcal{N}(0, \mathbf{I})} [\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|^2]. \quad (10)$$

To be more concise, \mathcal{L}_s refers to $\mathcal{L}_{\text{simple}}$ from Ho *et al.* [7].

3. Methodology

Natural images have been shown to have a power law distribution of frequencies [18, 25], where lower frequency components contribute more to the image. Due to Gaussian white noise having a uniform spectral density, low frequency components of partially diffused images do not become corrupted to the same extent as high frequency terms. This is readily apparent in Figure 1 where partial Gaussian diffusion leaves an easily identifiable view of the initial

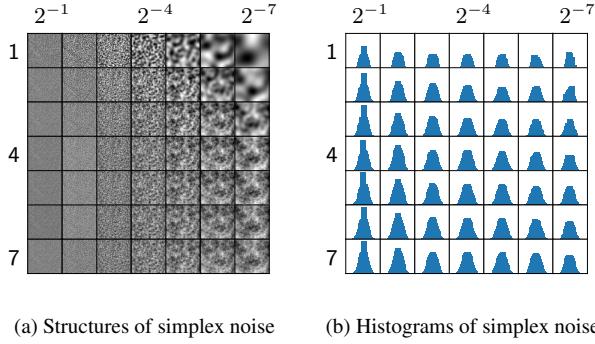


Figure 2. Simplex noise comparing the affect of the octave N (y-axis) and the initial starting frequency ν (x-axis) on the distribution and structure.

image. This limits the discriminatory power of an AnoDDPM model as low frequency components are inferred to be relatively corruption free, resulting in large anomalous regions being reconstructed in the reverse process. Working under the assumption that anomalous and non-anomalous images both follow power laws, we are motivated to modify the diffusion process so that the applied noise follows a similar power law, affecting low frequency components strongly. We wish to sample from a distribution which follows a power law. While these can be calculated through 2-dimensional Gaussian random fields or processes, or by careful construction of a co-variance matrix, we instead approximate it with simplex noise [13]. Simplex noise allows us to precisely control the distribution of frequencies present in the image. In contrast to classical white noise, simplex noise and related concepts produce *smooth* or structured randomness making them popular in computer graphics, for instance, for auto-generated structures such as landscapes.

3.1. Simplex Noise

For the 2-dimensional case, for example, Perlin noise is generated as follows: first, random gradients are sampled on a lattice. For any candidate point we then compute the inner product between the gradient and the offset of the candidate from the nearest 4 lattice points. The resulting values are subsequently interpolated, producing smooth noise. The more advanced simplex noise replaces the lattice with a simplex grid of equilateral triangles. This reduces the complexity with respect to the dimension and reduces directional artefacts over the typical gradient based Perlin noise. An example of simplex noise can be seen in Fig. 2. The structural difference is visually immediate and the potential benefit of such noise over the standard Gaussian perturbations is intuitive: the corruption is more structured and the denoising process will be able to “repair” those structured anomalies.

3.2. AnoDDPM

Our anomaly detection segmentation approach explores the use of DDPMs with either Gaussian or simplex noise. Initially, the query image x_0 is corrupted for t timesteps x_t which is then denoised back to \hat{x}_0 . As the anomaly can be of varying size, we parameterise x_t to x_λ , where larger λ values can remove larger anomalies.

Additionally, simplex noise has a set frequency, which we can adjust to allow a greater area of the image to become corrupted and reconstructed as healthy from the denoising process.

Instead of using the default simplex noise function, we can apply a number of octaves of noise (also known as fractal noise). This involves combining N frequencies of noise together, where the next frequency’s amplitude reduces by some decay rate γ . Figure 2b shows that low frequency noise cannot be well approximated with a Gaussian distribution; however, by applying an increasing number of octaves of noise, the distribution becomes closer to a Gaussian distribution. This is paramount for our DDPM model as we assume our noising function is sampling from a Gaussian distribution. Therefore, unless stated otherwise, we use a starting frequency of $\nu = 2^{-6}$, octave of $N = 6$ and a decay of $\gamma = 0.8$. Furthermore, when generating the simplex noise, we shuffle the seed before every noise calculation, and take a slice t from the 3-dimensional noise function, as we found that artefacts were introduced when sampling from the 2-dimensional noise function.

3.3. Training

We train our model only on healthy samples using the following training procedure:

Algorithm 1 Training

```

1: repeat
2:    $x_0 \sim q(x_0)$ 
3:    $t \sim \text{Uniform}(\{1, 2, \dots, T - 1, T\})$ 
4:   Randomly generate simplex seed
5:    $\epsilon \sim \text{Simplex}(\nu = 2^{-6}, N = 6, \gamma = 0.8)$ 
6:   Take gradient descent step on:
       $\nabla_\theta [\|\epsilon - \epsilon_\theta(x_0 \sqrt{\alpha_t} + \sqrt{1 - \alpha_t} \epsilon, t)\|^2]$ 
7: until converged

```

3.4. Inference

At inference, we take a query from our anomalous dataset A , then noise it to some timestep $x_0 \rightarrow x_\lambda$ where it is then denoised back to $x_\lambda \rightarrow x_0$. This can be seen as a conditional DDPM model on the input query (see Fig. 1). Unless stated otherwise, we use $\lambda = 250$ for our experiments. Therefore, the inference-time efficiency is $\mathcal{O}(\lambda)$. Following this reconstruction, we find the square error be-

tween the reconstruction and the initial image: $(x_0 - \hat{x}_0)^2$, which is shown as a heatmap throughout this paper. Then, to segment the tumour, we take a naïve threshold such that any error larger than 0.5 will be highlighted. For evaluation, we can then compare the prediction to the ground truth.

Algorithm 2 Segmentation

```

1:  $x_0 \sim A(x_0)$ 
2:  $x_\lambda = x_0\sqrt{\bar{\alpha}_\lambda} + \epsilon\sqrt{1 - \bar{\alpha}_\lambda}$ 
3: for  $t = \lambda, \dots, 1$  do
4:   Randomly generate simplex seed
5:    $z \sim \text{Simplex}(2^{-6}, 6, 0.8)$  if  $t > 0$  else  $z = 0$ 
6:    $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\epsilon_\theta(x_t, t)) + \tilde{\beta}_t z$ 
7: end for
8:  $E_{sq} = (x_0 - \hat{x}_0)^2$ 
9:  $E_{seg} = E_{sq} > 0.5$ 
10: return  $E_{seg}$ 
```

3.5. Implementation and Availability

For all our DDPM experiments, we use the same U-Net [17] architecture from Dhariwal and Nichol [5] for our ϵ_θ approximation. This is based on PixelCNN [19] and Wide ResNet [28], with transformer sinusoidal positional embedding [26] to encode the timestep. The model hyperparameters are:

Hyperparameters	
Diffusion steps	1000
Noise schedule	linear
Channels	128
Channels multiple	1, 1, 2, 3, 4
Heads	2
Attention resolution	32, 16, 8
Dropout	0
EMA rate	0.9999
Optimiser	AdamW [9]
Learning rate	1e-4
β_1, β_2	0.9, 0.999
Batch size	1
Slices trained	100 × 5000

The model is implemented in PyTorch and trained on a single NVIDIA Titan Xp GPU with 12GB GDDR5. The code is released under the MIT license and made available at: (Github URL <https://github.com/J Julian-Wyatt/AnoDDPM>).

3.6. Datasets

For our healthy dataset, we utilise the Neurofeedback Skull-Stripped (NFBS) repository [16] which contains 125 T1-weighted MRI scans with dimension $256 \times 256 \times 192$,

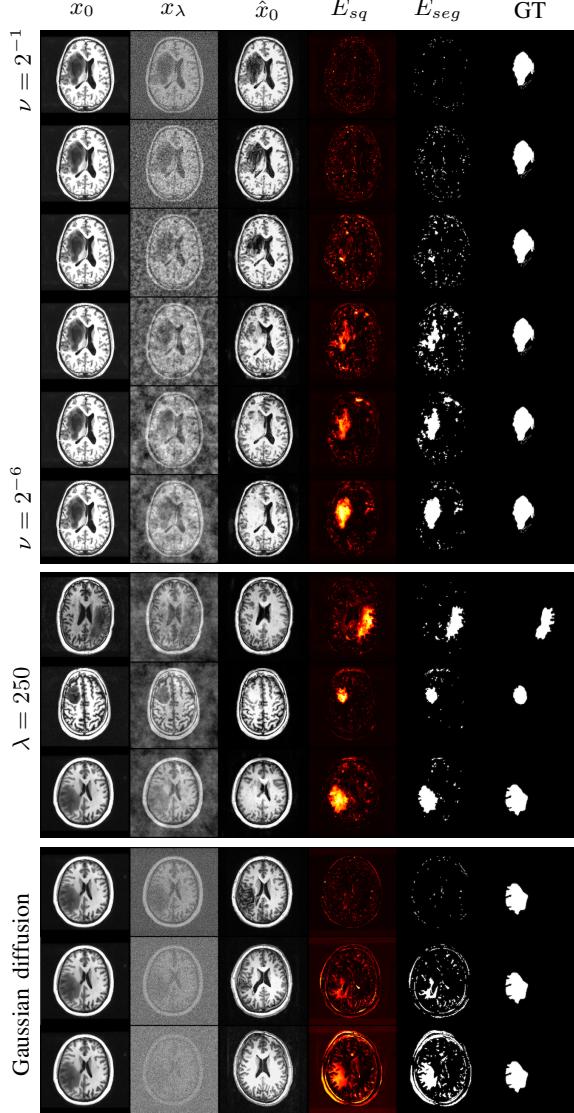


Figure 3. Qualitative comparison of simplex and Gaussian noise for detection, with x_0 , x_λ , \hat{x}_0 , square error ($E_{sq} = (x_0 - \hat{x}_0)^2$), prediction (E_{seg}) and ground truth (GT). Increasing multi-scale frequency of simplex noise with $\nu = 2^{-1}, \dots, 2^{-6}$ (top), curated samples (middle), Gaussian diffusion for $\lambda = 250, 500, 750$ (bottom).

containing the full skull, skull stripped and brain mask. We use the full skull images as an anomaly can appear anywhere, although this increases the complexity of the problem. We use 2D 256×192 axial slices of the brain as anomalies are generally easier to spot from this view. We then select the slice by randomly selecting an integer $i \in [40, 100]$. For the preprocessing, we apply a random rotation of $\pm 3^\circ$ and a random translation of $0.02 \times \text{width}$ and $0.09 \times \text{height}$, and then a center crop of 235, which is resized to 256×256 . This is split into 100 training and 25 testing volumes.

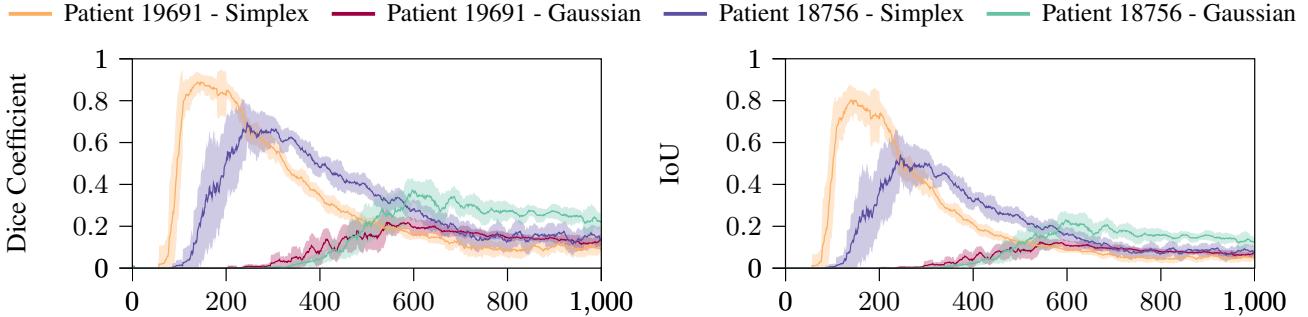


Figure 4. Segmentation evaluation experiment for patient 19691 at slice 171 and patient 18756 at slice 170 where $\lambda = 1, \dots, 1000$. (a) Dice Coefficient, (b) Intersection over Union

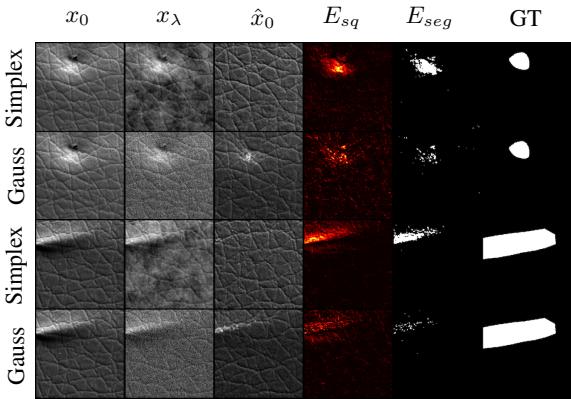


Figure 5. Preliminary qualitative comparison of simplex and Gaussian noise on the MVTec AD leather subset [2] for detection, with x_0 , x_λ , \hat{x}_0 , E_{sq} , E_{seg} , and ground truth (GT). We set $\lambda = 150$ for both samples, while the threshold value was moved from 0.5 to 0.15 for rows 1-2 and 0.2 for rows 3-4.

For evaluation, we use an anomalous dataset of brain tumours, unless stated otherwise. The dataset contains 22 T1-weighted MRI scans provided by the Centre for Clinical Brain Sciences from the University of Edinburgh [14]. The dataset consists of non-skull-stripped scans of $256 \times 256 \times 156$. This dataset contains image intensity non-uniformities that are the result of magnetic field variations rather than anatomical differences. Therefore we applied the BrainSuite [22] bias field correction to the images. For our analysis, we center crop the axial scans to 175×240 which are then resized to 256×256 in order to maintain the aspect ratio of the training set.

4. Experiments and Results

For our evaluation, we try to segment anomalies across all 22 images in the dataset [14], where we use 4 equally spaced slices that contain the tumour. We can perform the conditional reconstruction with Gaussian noise [7], or with

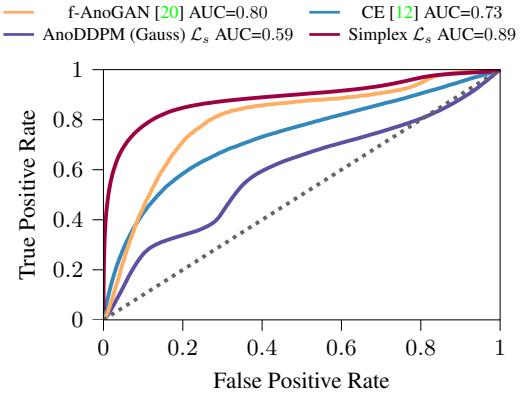


Figure 6. ROC curve comparing square error probabilities across every flattened anomalous MRI image [16]

simplex noise. Therefore, we evaluate the effect that the noise function has on anomaly detection.

Furthermore, we sample AnoDDPM with simplex and Gaussian noise on the leather subset of the MVTec AD dataset [2], which results in excellent reconstructions, as seen in Fig. 5; therefore, it can work well in alternative domains. We preprocess the dataset by converting the images to greyscale and generate a 256×256 random crop from the original 1024×1024 image.

4.1. Gaussian vs. simplex noise

As seen in Fig. 3 (bottom), it is clear that Gaussian noise produces higher quality samples, however, after corrupting the image to a larger t , the details of the image are lost and altered, where at $\lambda = 750$ a completely different sample is generated. Conversely, the simplex noise in Fig. 3 (top & middle) captures the effect that the frequency of noise has on anomaly detection. It shows that as the maximum frequency of multi-scale noise increases, more of the tumour is corrupted and then reconstructed as a healthy approximation.

We also evaluate the release time t that produces the best

	Dice \uparrow	IoU \uparrow	Precision \uparrow	Recall \uparrow	AUC \uparrow
Context Encoder [12]	0.252 ± 0.209	0.162 ± 0.149	0.258 ± 0.223	0.279 ± 0.234	0.707 ± 0.150
f-AnoGAN [20]	0.128 ± 0.001	0.093 ± 0.003	0.362 ± 0.009	0.080 ± 0.003	0.789 ± 0.001
AnoDDPM - Gauss (Ours)	0.009 ± 0.012	0.004 ± 0.006	0.006 ± 0.009	0.032 ± 0.044	0.601 ± 0.074
AnoDDPM \mathcal{L}_s (Ours)	0.383 ± 0.258	0.269 ± 0.204	0.373 ± 0.269	0.468 ± 0.283	0.863 ± 0.107

Table 1. Segmentation performance for a threshold of 0.5 on the anomalous dataset (where AUC uses the square error probability prediction averaged across each volume). The metrics evaluated for the Context Encoder [12], only consider the single centered 64x64 inpainted region in a 128x128 slice; while we calculated the f-AnoGAN [20] metrics using the square error of a 64x64 slice as this gave better results over absolute error proposed in [20].

segmentation, as seen in Fig. 4. The two large tumours result in a very different optimal release time t which shows that the optimal t depends on more than just the size of the anomalous region. Hence, performing a grid search, we found empirically that $\lambda = 250$ was optimal for this particular data set.

Lastly, we calculate the receiver operating characteristic curve (ROC curve), Fig. 6, which compares a context encoder reconstruction approach [12], f-AnoGAN [20], AnoDDPM with Gaussian noise [7], and AnoDDPM with simplex noise. We use the squared error probabilities $(x - \hat{x})^2$ to find the true positive and false positive rates. As seen in Fig. 6, AnoDDPM with Gaussian noise [7] performs slightly better than randomly for our anomalous dataset [16], while AnoDDPM with simplex noise performs very well, and generally performs better than f-AnoGAN [20] (trained on 64x64 slices).

4.2. Discussion

We observe a slight decrease in sample quality for simplex models, especially when noised to a further t value. This is likely due to the asymmetry in the simplex noise function; alternatives to this could be explored in future work through exploring a wider range of noising functions or designing custom multi-frequency noise functions. Furthermore, because the algorithm behaves stochastically, future work could explore sampling x_λ more than once and averaging the corresponding reconstructions \hat{x}_0 .

Additionally, the model struggles to segment tumours which are situated near other high frequency information such as around the brain’s gyri and sulci. Further work could be performed here to improve this by noising different sections of the image by different amounts using inpainting methods. We also experienced samples that introduced features of the brain such as the lateral ventricle into slices that should not contain those features. Therefore, with a larger training and testing corpus, a guided diffusion model that is conditioned on the slice could be evaluated and explored.

In the future, we would like to see if the AnoDDPM approach with simplex noise can leverage the additional information in 3D images or colour images and extend to other anomaly detection image datasets and applications, such as

in x-ray baggage security, biological images, and counterfeit products.

5. Conclusion

In conclusion, we found that AnoDDPM with simplex noise successfully captures large anomalous regions with stable training that does not require large datasets, avoiding the limitations common in GAN-based approaches [27]. The use of multiscalar (simplex) noise instead of Gaussian noise was found to offer significant improvement in terms of capturing larger anomaly shapes in both medical and non-medical applications; in the future it would be beneficial to further investigate how different noise types correspond to other anomaly detection applications using the AnoDDPM approach, such as in automatic threat detection and counterfeit product detection. In particular, AnoDDPM with simplex noise in the application of T1-weighted brain MRI anomaly detection was found to outperform its Gaussian noise counterpart as well as f-AnoGAN [20] in segmentation (DICE, IoU) and classification metrics (AUC). The proposed partial diffusion strategy ensures the method remains performant in high-resolution imagery (images throughout paper are best zoomed in). We expect this approach to also be successful in other applications where a full length Markov chain is not required, such as in image enhancement, semantic segmentation and filtering.

6. Compliance

The training data used in this study [16] was gathered and anonymised by the Nathan Kline Institute who support large scale open access health data. The data was acquired in compliance with the Health Insurance Portability and Accountability Act (HIPAA). The evaluation data [14] was anonymised and made open access by the Centre for Clinical Brain Sciences, University of Edinburgh. The patients were given an identifiable random code which the data collectors kept a record of.

References

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, 2017.

- tional conference on machine learning, pages 214–223. PMLR, 2017. 2
- [2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtac ad-a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 5
- [3] Nitish Bhatt, David Ramón Prados, Nedim Hodzic, Christos Karanassios, and HR Tizhoosh. Unsupervised detection of lung nodules in chest radiography using generative adversarial networks. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3842–3845. IEEE, 2021. 2
- [4] Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G. Willcocks. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1
- [5] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 2021. 2, 4
- [6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1
- [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1, 2, 5, 6
- [8] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 2
- [9] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 4
- [10] Bao Nguyen, Adam Feldman, Sarah Bethapudi, Andrew Jennings, and Chris G Willcocks. Unsupervised region-based anomaly detection in brain MRI with adversarial image inpainting. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1127–1131. IEEE, 2021. 2
- [11] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 2
- [12] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016. 5, 6
- [13] Ken Perlin. Improving noise. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 681–682, 2002. 2, 3
- [14] Cyril Pernet, Krzysztof Gorgolewski, and Whittle Ian. A neuroimaging dataset of brain tumour patients. 2016. 5, 6
- [15] Walter Hugo Lopez Pinaya, Petru-Daniel Tudosi, Robert Gray, Geraint Rees, Parashkev Nachev, Sébastien Ourselin, and M Jorge Cardoso. Unsupervised brain anomaly detection and segmentation with transformers. *arXiv preprint arXiv:2102.11650*, 2021. 1
- [16] Benjamin Puccio, James P Pooley, John S Pellman, Elise C Taverna, and R Cameron Craddock. The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data. *GigaScience*, 5(1), 10 2016. s13742-016-0150-5. 4, 5, 6
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 4
- [18] Daniel L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3398, 1997. 2
- [19] Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017. 4
- [20] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 2, 5, 6
- [21] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017. 1
- [22] David W Shattuck and Richard M Leahy. Brainsuite: an automated cortical surface identification tool. *Medical image analysis*, 6(2):129–142, 2002. 5
- [23] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 2
- [24] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 1
- [25] A. van der Schaaf and J.H. van Hateren. Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17):2759–2770, 1996. 2
- [26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4
- [27] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion GANs. In *International Conference on Learning Representations*, 2022. 1, 2, 6
- [28] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016. 4
- [29] David Zimmerer, Simon AA Kohl, Jens Petersen, Fabian Isensee, and Klaus H Maier-Hein. Context-encoding variational autoencoder for unsupervised anomaly detection. *arXiv preprint arXiv:1812.05941*, 2018. 2