



HAL
open science

Unsupervised anomaly detection using Bayesian flow networks: application to brain FDG pet in the context of Alzheimer's disease

Hugues Roy, Reuben Dorent, Ninon Burgos

► To cite this version:

Hugues Roy, Reuben Dorent, Ninon Burgos. Unsupervised anomaly detection using Bayesian flow networks: application to brain FDG pet in the context of Alzheimer's disease. DGM4MICCAI 2025 - Deep Generative Models, Sep 2025, Daejong, South Korea. pp.254-264, <10.1007/978-3-032-05472-2_25>. <hal-05322316>

HAL Id: hal-05322316

<https://hal.science/hal-05322316v1>

Submitted on 20 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Unsupervised anomaly detection using Bayesian flow networks: application to brain FDG PET in the context of Alzheimer’s disease

Hugues Roy¹, Reuben Dorent^{1,2}, and Ninon Burgos¹

¹ Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

² MIND Team, Inria Saclay, Université Paris-Saclay, Palaiseau, France

Abstract. Unsupervised anomaly detection (UAD) plays a crucial role in neuroimaging for identifying deviations from healthy subject data and thus facilitating the diagnosis of neurological disorders. In this work, we focus on Bayesian flow networks (BFNs), a novel class of generative models, which have not yet been applied to medical imaging or anomaly detection. BFNs combine the strength of diffusion frameworks and Bayesian inference. We introduce AnoBFN, an extension of BFNs for UAD, designed to: i) perform conditional image generation under high levels of spatially correlated noise, and ii) preserve subject specificity by incorporating a recursive feedback from the input image throughout the generative process. We evaluate AnoBFN on the challenging task of Alzheimer’s disease-related anomaly detection in FDG PET images. Our approach outperforms other state-of-the-art methods based on VAEs (β -VAE), GANs (f-AnoGAN), and diffusion models (AnoDDPM), demonstrating its effectiveness at detecting anomalies while reducing false positive rates.

Keywords: Generative models · Bayesian Flow Networks · Anomaly detection · Neuroimaging

1 Introduction

Pixel-wise anomaly detection is a key problem in medical image computing. The current state-of-the-art approach is supervised learning (segmentation), which requires large and manually annotated datasets. Alternatively, unsupervised anomaly detection (UAD) has raised significant interest as it bypasses the need for annotation by training a generative model on healthy data [4]. At inference, the model reconstructs a pseudo-healthy version of a given image, and anomalies are identified by comparing the original and reconstructed images, enabling the detection of lesions without prior knowledge of their specific appearance.

Classical UAD generative models include the f-AnoGAN [18] and variational autoencoders (VAEs) [11,2,7]. However, these models, trained only on healthy data, are sensitive to distribution shifts and often produce unreliable encodings for anomalous inputs [3]. Consequently, normal regions may be misreconstructed, leading to a loss of subject specificity and a large rate of false positives.

Denosing diffusion models [9,19,10] address the encoding limitation by using a latent representation based on noisy versions of the data. These models have demonstrated competitive performance in chest x-ray [23] or brain MRI [14,23,25,26] applications. A key limitation of diffusion models is their tendency to change regions even when no anomalies are present. Higher noise levels are required to eliminate larger anomalies but reduce subject’s specificity. Therefore, an optimal balance must be achieved between effective anomaly removal and preservation of subject’s specificity.

In this context, we leverage the use of Bayesian flow networks (BFNs) [5], a novel class of generative models based on diffusion models, introducing a Bayesian latent representation on the parameters of a data distribution rather than on the data directly. The latent parameter variable allows information aggregation of past generated samples and uses Bayesian inference to update the current state. BFNs have shown promise in 3D molecule modeling [1], material generation [24], and crystal generation [17] but, to our knowledge, have not yet been used in medical imaging or for anomaly detection.

We present AnoBFN, the first use of BFNs for unsupervised anomaly detection. First, we adapt BFNs to anomaly detection, introducing a novel probabilistic framework for this task. Second, we use simplex noise with a new accuracy schedule for conditional generation, preserving subject specificity under high spatially correlated noise. Third, we introduce a recursive Bayesian update scheme that fuses the original image with the model’s prediction to retain input information. Finally, we validate our approach on FDG-PET scans for Alzheimer’s disease detection, a clinically relevant task characterized by diffuse metabolic anomalies and absence of ground truth. To enable quantitative evaluation, we generate synthetic anomalies that replicate disease-like abnormalities and compare our method against state-of-the-art unsupervised anomaly detection models.

2 Bayesian flow networks

Bayesian flow networks (BFNs) [5] extend denosing diffusion models (DDMs) [9] by operating on the parameters of data distributions rather than directly on noisy data samples. BFNs aim to learn the parameters θ_0 of an *input distribution* that approximates the true data distribution p_{data} . This is done by an iterative process of steps $t = T, \dots, 0$ that refines the parameters $\theta_t = \{\mu_t, \rho_t\}$, considered as latent variables, of the input distribution, where $p_I(\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_t; \mu_t, \rho_t^{-1} \mathbf{I})$ and \mathbf{x}_t denotes the noisy data at time t . To update θ_t , the authors propose a principled formulation based on Bayes’ rule, leveraging observed \mathbf{x}_t or estimated $\hat{\mathbf{x}}_t$ data. The original BFN framework is illustrated in Fig. 1. During training, the observed data \mathbf{x}_t is drawn from a *sender distribution* $p_S(\mathbf{x}_t | \mathbf{x}_0, \alpha_t) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0, \alpha_t^{-1} \mathbf{I})$ by adding noise to the true data \mathbf{x}_0 according to a predefined accuracy parameter $\alpha_t = \frac{d\beta(t)}{dt}$ derived from the *accuracy schedule* β_t , as in DDMs. In the original framework, the noise level initially with t , then decreases, reaching a minimum at $t = 0$, where the distribution closely approximates true

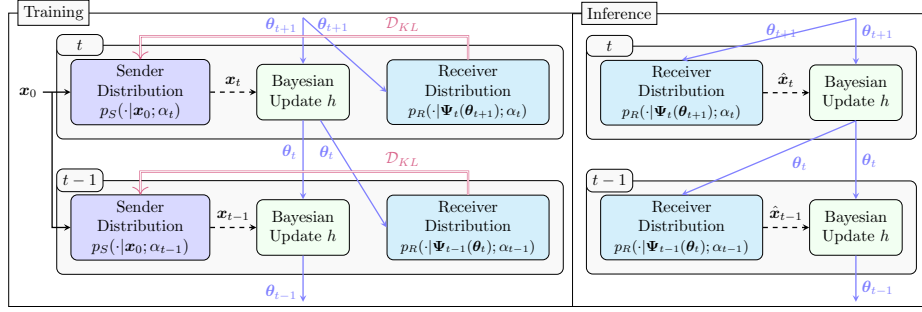


Fig. 1. Training and inference phases of the original Bayesian flow networks. Dashed lines correspond to samples from distributions, notations are defined in the main text.

data. This design enables sampling from p_{data} by initializing the generation process from a zero-mean prior. Then, Bayesian inference is used to update the parameters θ_t by sampling from the following *update distribution* p_U

$$p_U(\theta_t \mid \theta_{t+1}; \mathbf{x}_0, \alpha_t) = \mathbb{E}_{p_S(\mathbf{x}_t | \mathbf{x}_0; \alpha_t)} [\delta(\theta_t - h(\theta_{t+1}, \mathbf{x}_t, \alpha_t))] \quad (1)$$

where δ is the Dirac delta function and h the Bayesian update function computing the posterior using Bayes' rule and a Gaussian conjugate prior θ_{t+1}

$$h : \theta_{t+1}, \mathbf{x}_t, \alpha_t \mapsto \left(\boldsymbol{\mu}_t = \frac{\rho_{t+1} \boldsymbol{\mu}_{t+1} + \alpha_t \mathbf{x}_t}{\rho_{t+1} + \alpha_t}, \rho_t = \rho_{t+1} + \alpha_t \right) \quad (2)$$

θ_t represents the updated belief about the data after observing \mathbf{x}_t . This process mimics Bayesian inference, gradually refining estimates as evidence is gathered. At inference time, the estimated data $\hat{\mathbf{x}}_t$ is sampled from a *receiver distribution*, which follows the same noise schedule α_t . Specifically, a neural network Ψ predicts the mean of the receiver distribution using the current parameters θ_{t+1} , which are either sampled from the prior distribution p_P (if $t + 1 = T$) or the update Bayesian distribution p_U . The receiver distribution is defined as

$$p_R(\hat{\mathbf{x}}_t \mid \theta_{t+1}; \alpha_t) = \mathcal{N}(\hat{\mathbf{x}}_t; \Psi(\theta_{t+1}), \alpha_t^{-1} \mathbf{I}) \quad (3)$$

Similarly to Eq. (1), the parameters θ_t used at inference are sampled from the update distribution p_U , determined by the previous parameters θ_{t+1} , the estimated data $\hat{\mathbf{x}}_t$ sampled from the receiver distribution p_R defined in Eq. (3) and the Bayesian update h

$$p_U(\theta_t \mid \theta_{t+1}; \alpha_t) = \mathbb{E}_{p_R(\hat{\mathbf{x}}_t | \theta_{t+1}; \alpha_t)} \delta(\theta_t - h(\theta_{t+1}, \hat{\mathbf{x}}_t, \alpha_t)) \quad (4)$$

To ensure that the estimated data $\hat{\mathbf{x}}_t$ from the receiver distribution are faithful, the Kullback-Leibler divergence (D_{KL}) between the sender p_S and receiver p_R distributions is minimized during training across different timesteps and data

$$\mathcal{L}(\Psi) = \mathbb{E}_{\mathbf{x}_0 \sim p_{\text{data}}} \mathbb{E}_{p_U(\theta_t | \theta_T, \mathbf{x}_0, \beta_{t+1})} D_{KL} [p_S(\cdot | \mathbf{x}_0, \alpha_t) \parallel p_R(\cdot | \theta_{t+1}; \alpha_t)] \quad (5)$$

While BFNs have originally been introduced for generation purposes [5], they have not been adapted to anomaly detection yet.

3 Methods

In this work, we propose AnoBFN, an approach using Bayesian flow networks for unsupervised anomaly detection in images. Following the standard UAD framework, we first train a generative model to approximate the distribution of healthy data p_H , with samples $\mathbf{x}_{0,H}$. Then, we leverage this trained model to generate healthy versions of images $\mathbf{x}_{0,A}$ drawn from an abnormal distribution p_A , enabling anomaly detection. To achieve this objective, we extend the original BFN method to perform conditional image generation in the presence of anomalies.

Specifically, our contributions focus on two key objectives: [C1] enabling conditional generation under high levels of spatially correlated noise, which are needed to encourage overlapping prior distributions between abnormal scans and their pseudo-healthy counterparts; [C2] preserving subject specificity by incorporating recursive feedback from the input throughout the generative process.

For our first contribution [C1], we **jointly use simplex noise** and introduce a **new accuracy schedule**.

[C1.1] *Structured noise.* Gaussian noise is independent and identically distributed, lacking spatial coherence with a flat power spectrum in the Fourier decomposition space and high-frequency spatial variations [25]. In contrast, simplex noise provides gradient-based noise with spatial continuity and structured perturbations. Inspired by [25], we use structured noise to encourage the denoiser to handle spatially correlated perturbations rather than independent pixel noise.

[C1.2] *Accuracy schedule.* To reconstruct a pseudo-healthy version $\mathbf{x}_{0,H}$ from an abnormal input $\mathbf{x}_{0,A}$, the generative process should (i) be conditional and (ii) operate under a high-noise regime to ensure overlap between the prior distributions of healthy and abnormal data. In practice, the mean parameters $\boldsymbol{\mu}_t$ follow stochastic trajectories defined by the *probability flow distribution* p_F with mean and variance characterized by the accuracy schedule $\beta(t)$:

$$p_F(\boldsymbol{\mu}_t | \mathbf{x}_0; t) = p_U(\boldsymbol{\mu}_t | \boldsymbol{\theta}_T, \mathbf{x}_0, \beta(t)) \quad (6)$$

$$= \mathcal{N}\left(\boldsymbol{\mu}_t; \frac{\beta(t)\mathbf{x}_0 + \rho_T\boldsymbol{\mu}_T}{\rho_T + \beta(t)}, \frac{\beta(t)}{(\rho_T + \beta(t))^2}\mathbf{I}\right). \quad (7)$$

Since the original BFN is designed for unconditional generation, the prior distribution of $\boldsymbol{\mu}_T$ does not exploit subject-specific information, i.e. $\beta(T) = 0$ (Fig. 2). In contrast, our goal is to design a scheduler β such that: i) the prior distribution depends on the input, i.e. $\beta(T) > 0$; ii) the abnormal and normal prior distributions overlap, i.e. the variance of the prior of $\boldsymbol{\mu}_T$ is large; and iii) it is well-defined for Bayesian update. Under the prior $\boldsymbol{\theta}_T = \{\boldsymbol{\mu}_T, \rho_T\} = \{\mathbf{0}, 1\}$, the condition iii) can be satisfied if there is a function $f \in C([0, T])$ such that $f([0, T]) \subset (0, \frac{1}{4}]$ and for all $t \in [0, T]$, $f(t) = \frac{\beta(t)}{(1+\beta(t))^2}$. Inspired by [13], a suitable choice shown in Fig. 2 with an initial high variance decreasing over time can be expressed as:

$$f(t) = \frac{1}{4} \cos^4\left(\frac{(T-t) + s\pi}{T(1+s)}\right) \quad (8)$$

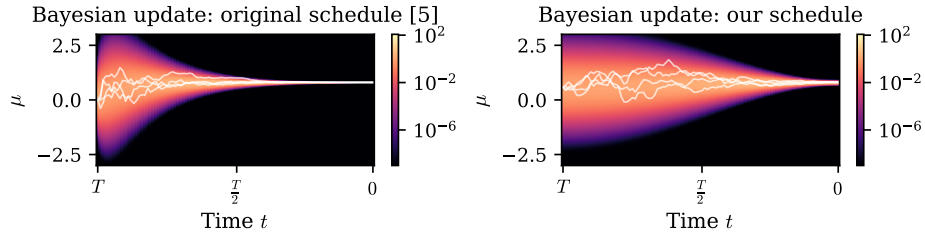


Fig. 2. The white curves illustrate sample trajectories of μ , the mean parameters of the latent variable θ , under both the schedule from [5] and our accuracy-based schedule. The background shows the probability flow distribution.

with $s = 0.01$. This introduced scheduler i) preserves information about the input \mathbf{x}_0 at $t = T$, as the initial mean of the probability flow distribution p_F is approximately $\frac{\mathbf{x}_0}{2}$ and ii) has maximal variance at $t = T$.

Overall, the combination of simplex noise and the modified probability flow enables conditional generation under high spatially correlated noise, encouraging overlapping prior distributions of μ_T for normal and abnormal scans.

[C2] *Bayesian Update.* Our second key objective is to preserve the subject specificity throughout the generative process at inference time. In classical anomaly detection frameworks, the abnormal image is only used once as input and to compute the residual, i.e., the difference between the input and reconstruction. In this work, we leverage the Bayesian update of BFNs to leverage the input image throughout the generation process. Specifically, we propose a novel Bayesian update that incorporates both the estimated noisy sample $\hat{\mathbf{x}}_t$ drawn from the learned receiver distribution $p_R(\hat{\mathbf{x}}_t | \theta_{t+1}; \alpha_t) = \mathcal{N}(\hat{\mathbf{x}}_t; \Psi(\theta_{t+1}), \alpha_t^{-1} \mathbf{I})$ and the input image \mathbf{x}_0 , allowing the latent variable to retain information from both sources. Therefore, the update distribution also depends on \mathbf{x}_0

$$p_U(\theta_t | \theta_{t+1}, \hat{\mathbf{x}}_t, \mathbf{x}_0) . \quad (9)$$

Similarly to the original framework, the Bayesian update employs a matrix α_t to weight the contribution of $\hat{\mathbf{x}}_t$. To additionally weight the impact of the observed input image \mathbf{x}_0 in the generative process during inference, we introduce an auxiliary weighting parameter $\alpha_{t,A}$ that modulates its contribution. The Bayesian update function is then defined as $h : \theta_{t+1}, \mathbf{x}_t, \alpha_t, \mathbf{x}_0, \alpha_{t,A} \mapsto (\mu_t, \rho_t)$ such that:

$$\begin{cases} \mu_t = \frac{1}{\rho_t} (\rho_{t+1} \odot \mu_{t+1} + \alpha_t \odot \hat{\mathbf{x}}_t + \alpha_{t,A} \odot \mathbf{x}_0) \\ \rho_t = \rho_{t+1} + \alpha_t + \alpha_{t,A} \end{cases} \quad (10)$$

Since the parameter $\alpha_{t,A}$ weights the contribution of the input image, $\alpha_{t,A}$ is expected to be larger in normal regions than in abnormal ones. To quantify in an unsupervised manner regions that are likely to be abnormal during the inference process, we postulate that abnormal regions are the most varying regions in the generation process. For this reason, we modulate the parameter $\alpha_{t,A}$ relying on the predefined accuracy parameter α_t using a scaling metric based on the

squared error between the generated mean of the receiver distribution and the input scan, and a time scaling factor ensuring the decay of $\alpha_{t,A}$. The higher the squared error, the less the update accounts for the input. For the time scaling, we decided to use the logistic function, with a logistic growth rate $k = 30$ and $t_c = 0.5$, such that the value of the function’s midpoint is at the middle of the generative process

$$\alpha_{t,A} = \alpha_t \odot e^{-(\psi(\theta_{t+1}) - \mathbf{x}_0)^2} \left(\frac{1}{1 + e^{-k(\frac{t}{T} - t_c)}} \right). \quad (11)$$

This formulation ensures that the Bayesian update accounts for structural differences between normal and anomalous regions while leveraging prior knowledge from the generative process.

4 Experiments and results

In this section, we demonstrate the effectiveness of our approach in detecting anomalies in FDG PET scans associated with Alzheimer’s disease (AD). A key challenge when evaluating UAD in such context is the absence of ground truth masks of the anomalies, contrary to applications that aim to detect lesions on structural MRI [2]. To address this, we assess our method using abnormal data obtained by simulating realistic hypometabolism characteristic of AD [6].

Dataset. The FDG PET scans used in this study were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database [21,22]. We downloaded the scans which frames were co-registered, averaged, and standardized to a uniform resolution. Subsequent image processing was conducted using Clinica [16], which included linear registration to the MNI space, intensity normalization using the mean PET uptake within the cerebellum and pons, and cropping. The images were further resampled to a grid size of $128 \times 128 \times 128$ voxels and rescaled to the range $[-1, 1]$. From each volume, 20 axial slices centered around the central slice were extracted. We selected 733 scans from 301 cognitively normal (CN) subjects. First, 80 subjects/scans were isolated as test set. A synthetic image mimicking 30% AD-induced hypometabolism was generated for each test scan using the method described in [6]. The healthy images, denoted by test_{CN} , were used to evaluate reconstruction performance while their synthetic abnormal counterparts, denoted by test_{sAD} , were used to assess anomaly detection performance. The remaining scans were split into training and validation sets at the subject level using ClinicaDL [20], with stratification based on age and sex, resulting in a training and validation set of 540 and 57 scans, respectively.

Baselines, implementation details and evaluation metrics. We compare our results with the β -VAE [8], the f-AnoGAN [18], and the AnoDPPM [25], so each generative model class is represented in our experiments. We used the same U-net [15] architecture from [9] for AnoDDPM and AnoBFN. The structure was modified resulting in an encoder (decoder) composed of 3 downsampling (upsampling) stages with layer width $[C, C, 2C]$ with $C = 128$. The attention layers are composed of 4 heads with dimensions 16. The optimizer was AdamW [12]

Table 1. Pseudo-healthy reconstruction and anomaly detection performance metrics. Standard deviations are calculated at the subject level. Statistical significance between AnoBFN and alternative methods was determined using a Bonferroni-corrected paired Wilcoxon signed-rank test, * indicating statistically significant differences ($p < 0.01$).

		Pseudo-healthy reconstruction (test_{CN})			Anomaly detection (test_{sAD})	
		MSE (10^{-3}) ↓	PSNR ↑	SSIM ↑	IoU ↑	AP ↑
β -VAE	[8]	$17.02 \pm 3.21^*$	$23.84 \pm 0.83^*$	$69.98 \pm 3.57^*$	$22.82 \pm 5.31^*$	$28.05 \pm 8.43^*$
f-AnoGAN	[18]	$31.70 \pm 8.87^*$	$21.23 \pm 1.11^*$	$59.78 \pm 5.63^*$	$17.56 \pm 4.65^*$	$21.36 \pm 7.82^*$
AnoDDPM (simplex)	[25]	8.72 ± 1.33	26.81 ± 0.60	$79.79 \pm 2.75^*$	$15.95 \pm 4.77^*$	$26.41 \pm 7.21^*$
BFN	[5]	$38.48 \pm 4.43^*$	$20.30 \pm 0.47^*$	$54.38 \pm 2.69^*$	$8.46 \pm 1.68^*$	$11.27 \pm 2.29^*$
AnoBFN w/o [C2]		$23.53 \pm 3.39^*$	$22.45 \pm 0.59^*$	$66.07 \pm 3.68^*$	$19.99 \pm 3.82^*$	$28.95 \pm 6.88^*$
AnoBFN (Ours)		9.67 ± 1.66	26.28 ± 0.72	80.92 ± 2.62	31.79 ± 7.74	48.17 ± 11.31

with learning rate $1e - 4$, weight decay 0.01 and $(\beta_1, \beta_2) = (0.9, 0.98)$. An exponential moving average of model parameters with a decay rate of 0.9999 was used for evaluation and sample generation for AnoDDPM and AnoBFN. The total number of learnable parameters was $\sim 10\text{M}$. The batch size was 30 and we used gradient clipping to 1 during the training steps. Regarding the β -VAE, we use $\beta = 10$, a latent dimension of size 128, and a symmetrical encoder-decoder architecture using batch normalization and LeakyRelu activation functions with channel multiplication [1,2,6,8]. For the f-AnoGAN, we used the same architecture as the β -VAE. The code is available at <https://github.com/HuguesRoy/BFNanom>.

We evaluate reconstruction quality using the mean square error (MSE), peak-to-signal noise ratio (PSNR), and structural similarity (SSIM) computed for healthy scans of the test set test_{CN} . We assess anomaly detection performance using simulated abnormal scans from test_{sAD} by computing the intersection over union (IoU) for a threshold of 0.05 corresponding to the mean square synthetic mask’s intensity decrease, and the average precision (AP) at the pixel-level [4].

Results. Before evaluating performance on the primary task of anomaly detection, we first assess reconstruction performance on healthy data as a form of sanity check. While achieving high-quality reconstructions is not the primary objective, this step helps ensure that the models have learned meaningful representations of normal anatomy. As shown in Table 1, we can observe that both f-AnoGAN and β -VAE poorly reconstruct healthy images and underperform compared to AnoDDPM and AnoBFN, both producing similar results. AnoDDPM and f-AnoGAN are the methods the least able to detect anomalies, followed by β -VAE and AnoBFN. AnoBFN outperforms all models in AP and IoU ($p < 0.01$). We also conducted an ablation study to assess each of our contributions to the original BFN. We can observe that introducing simplex noise and accuracy schedule [C1] (AnoBFN w/o [C2] vs BFN) leads to significant improvements in both healthy reconstruction and anomaly detection. Adding the Bayesian update guidance [C2] (AnoBFN) further improves our model, leading to state-of-the-art performance. Qualitative results are shown in Fig. 3. The β -VAE produces a blurred image, struggling to preserve high-uptake regions

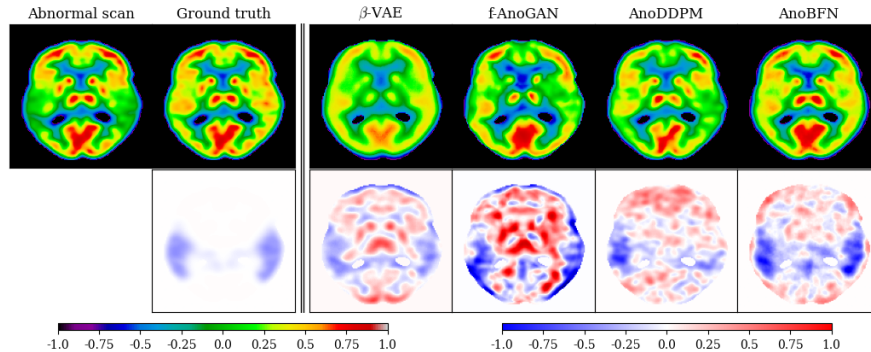


Fig. 3. Example of reconstructions and residual maps for a random subject in the test set. Top: synthetic abnormal scan, ground truth (real scan without anomalies) and pseudo-healthy reconstructions generated by the different models from the synthetic abnormal scan. Bottom row: mask used to simulate hypometabolism and difference maps computed between each reconstruction and the abnormal scan.

and fine details. f-AnoGAN, while not producing an exact reconstruction, effectively preserves the overall shape and shows higher contrast between normal and anomalous regions. AnoDDPM closely matches the original abnormal scan, preserves its structure, and captures a substantial portion of the anomaly. Finally, AnoBFN produces a sharp reconstruction that preserves subject’s specificity while restoring healthy-appearing tissue in originally abnormal regions, leading to a clear delineation of anomalies.

5 Conclusion and future works

We introduce Bayesian flow networks [5] to medical imaging for the first time, demonstrating their ability to perform unsupervised anomaly detection. To achieve this, we extended the original BFN framework to enable effective pseudo-healthy reconstruction and preservation of the subjects’ identity. AnoBFN introduces two key contributions: i) the combination of simplex noise and an accuracy schedule enabling conditional generation under high levels of spatially correlated noise, ii) a novel Bayesian update mechanism that recursively incorporates input information during inference, preserving subject specificity. Our approach outperforms state-of-the-art methods [8,18,23] in the detection of synthetic anomalies simulating Alzheimer’s disease in FDG PET imaging. We confirmed the key role of the proposed Bayesian update by performing an ablation study. To enhance the robustness of anomaly detection, we aim to incorporate uncertainty quantification into the Bayesian generative process while also exploring alternative accuracy schedules and scaling metrics for the Bayesian update. Beyond methodological refinements, we aim to evaluate the robustness and generalization of our approach by applying it to a broader range of medical imaging datasets, including diverse pathologies and imaging modalities.

6 Acknowledgement

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program (references ANR-19-P3IA-0001 - PRAIRIE - and reference ANR-10-IAHU-0006) and the “France 2030” program (reference ANR-23-IACL-0008, PRAIRIE-PSAI), and from the European Union’s Horizon Europe Framework Program (grant number 101136607, project CLARA). This research is part of the MediTwin project, which has been funded by the French government as part of “France 2030” and by the European Union - Next Generation EU as part of the “France Relance” plan. This work was performed using HPC resources from GENCI-IDRIS (Grant 2025-AD011011648R5). R.D. received a Marie Skłodowska-Curie grant No 101154248 (project: SafeREG). Data collection and sharing for the Alzheimer’s Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH).

References

1. Atkinson, T., Barrett, T.D., Cameron, S., Guloglu, B., Greenig, M., Robinson, L., Graves, A., Copoiu, L., Laterre, A.: Protein Sequence Modelling with Bayesian Flow Networks (2024). <https://doi.org/10.1101/2024.09.24.614734>
2. 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* p. 101952 (2021). <https://doi.org/10.1016/j.media.2020.101952>
3. Cai, Y., Chen, H., Cheng, K.T.: Rethinking autoencoders for medical anomaly detection from a theoretical perspective. In: *Medical Image Computing and Computer-Assisted Intervention*. pp. 544–554 (2024). https://doi.org/10.1007/978-3-031-72120-5_51
4. Cai, Y., Zhang, W., Chen, H., Cheng, K.T.: MedIAnomaly: A comparative study of anomaly detection in medical images. *Medical Image Analysis* **102**, 103500 (2025). <https://doi.org/10.1016/j.media.2025.103500>
5. Graves, A., Srivastava, R.K., Atkinson, T., Gomez, F.: Bayesian Flow Networks (2024), <http://arxiv.org/abs/2308.07037>
6. Hassanally, R., Brianceau, C., Solal, M., Colliot, O., Burgos, N.: Evaluation of pseudo-healthy image reconstruction for anomaly detection with deep generative models: Application to brain FDG PET. *Machine Learning for Biomedical Imaging 2*(Special Issue for Generative Models), 611–656 (2024). <https://doi.org/10.59275/j.melba.2024-b87a>
7. Hassanally, R., Solal, M., Colliot, O., Burgos, N.: Pseudo-healthy image reconstruction with variational autoencoders for anomaly detection: A benchmark on 3D brain FDG PET (2024), <https://inria.hal.science/hal-04445378>

8. Higgins, I., Matthey, L., Pal, A., Burgess, C.P., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In: International Conference on Learning Representations (2017), <https://openreview.net/forum?id=Sy2fzU9gl>
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
10. Kingma, D., Salimans, T., Poole, B., Ho, J.: Variational diffusion models. In: Advances in Neural Information Processing Systems. vol. 34, pp. 21696–21707 (2021), https://proceedings.neurips.cc/paper_files/paper/2021/file/b578f2a52a0229873fefc2a4b06377fa-Paper.pdf
11. Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes (2022). <https://doi.org/10.48550/arXiv.1312.6114>
12. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>
13. Nichol, A., Dhariwal, P.: Improved Denoising Diffusion Probabilistic Models (Feb 2021). <https://doi.org/10.48550/arXiv.2102.09672>, <http://arxiv.org/abs/2102.09672>, arXiv:2102.09672 [cs]
14. Pinaya, W.H., Graham, M.S., Gray, R., Da Costa, P.F., Tudosiu, P.D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., et al.: Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: Medical Image Computing and Computer-Assisted Intervention. pp. 705–714 (2022). https://doi.org/10.1007/978-3-031-16452-1_67
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.O., Durrleman, S., Colliot, O.: Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies. *Frontiers in Neuroinformatics* **15**, 39 (2021). <https://doi.org/10.3389/fninf.2021.689675>
17. Ruple, L., Torresi, L., Schopmans, H., Friederich, P.: Symmetry-aware bayesian flow networks for crystal generation (2025), <https://arxiv.org/abs/2502.03146>
18. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis* **54**, 30–44 (2019). <https://doi.org/10.1016/j.media.2019.01.010>
19. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021), <https://openreview.net/forum?id=PXTIG12RRHS>
20. Thibeau-Sutre, E., Díaz, M., Hassanaly, R., Routier, A., Dormont, D., Colliot, O., Burgos, N.: ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing. *Computer Methods and Programs in Biomedicine* **220**, 106818 (2022). <https://doi.org/10.1016/j.cmpb.2022.106818>

21. Weiner, M.W., Aisen, P.S., Jack, C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., Toga, A., Green, R., Walter, S., Soares, H., Snyder, P., Siemers, E., Potter, W., Cole, P.E., Schmidt, M.: The Alzheimer’s Disease Neuroimaging Initiative: Progress report and future plans. *Alzheimer’s & Dementia* **6**(3), 202 (2010)
22. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Morris, J.C., Petersen, R.C., Salazar, J., Saykin, A.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q.: The Alzheimer’s Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer’s & Dementia* **13**(5), 561–571 (2017)
23. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: *Medical Image Computing and Computer-Assisted Intervention*. pp. 35–45 (2022). https://doi.org/10.1007/978-3-031-16452-1_4
24. Wu, H., Song, Y., Gong, J., Cao, Z., Ouyang, Y., Zhang, J., Zhou, H., Ma, W.Y., Liu, J.: A periodic bayesian flow for material generation (2025), <https://arxiv.org/abs/2502.02016>
25. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 649–655 (2022). <https://doi.org/10.1109/CVPRW56347.2022.00080>
26. Yu, J., Oh, H., Yang, J.: Adversarial denoising diffusion model for unsupervised anomaly detection. In: *Deep Generative Models for Health Workshop NeurIPS 2023* (2023), <https://openreview.net/forum?id=PvHhnn1iX9>