



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

# Differentiable Score-Based Likelihoods: Learning CT Motion Compensation From Clean Images

Mareike Thies<sup>1</sup>[0000-0002-1364-4337], Noah Maul<sup>1</sup>[0000-0003-1161-1299],  
Siyuan Mei<sup>1</sup>[0009-0002-1136-2501], Laura Pfaff<sup>1</sup>,  
Nastassia Vysotskaya<sup>1</sup>[0000-0003-1367-7654], Mingxuan Gu<sup>1</sup>[0000-0002-5244-4397],  
Jonas Utz<sup>1</sup>, Dennis Possart<sup>1,2</sup>, Lukas Folle<sup>1</sup>[0000-0002-5395-097X],  
Fabian Wagner<sup>1</sup>[0000-0002-5964-6485], and Andreas Maier<sup>1</sup>[0000-0002-9550-5284]

<sup>1</sup> Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

<sup>2</sup> Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Forchheim, Germany

mareike.thies@fau.de

**Abstract.** Motion artifacts can compromise the diagnostic value of computed tomography (CT) images. Motion correction approaches require a per-scan estimation of patient-specific motion patterns. In this work, we train a score-based model to act as a probability density estimator for clean head CT images. Given the trained model, we quantify the deviation of a given motion-affected CT image from the ideal distribution through likelihood computation. We demonstrate that the likelihood can be utilized as a surrogate metric for motion artifact severity in the CT image facilitating the application of an iterative, gradient-based motion compensation algorithm. By optimizing the underlying motion parameters to maximize likelihood, our method effectively reduces motion artifacts, bringing the image closer to the distribution of motion-free scans. Our approach achieves comparable performance to state-of-the-art methods while eliminating the need for a representative data set of motion-affected samples. This is particularly advantageous in real-world applications, where patient motion patterns may exhibit unforeseen variability, ensuring robustness without implicit assumptions about recoverable motion types.

**Keywords:** Diffusion models · Neural ordinary differential equations · Exact likelihood computation · Computed tomography · Motion compensation.

## 1 Introduction

Spotting motion artifacts in CT images is an easy task. Even an inexperienced CT reader can differentiate an artifact-free from a motion-affected image. As observers of the images, we have an implicit understanding of a “good” CT image. Therefore, we can recognize deviations from that state as, e.g., introduced through motion. In contrast, neural networks usually need to be presented with examples of all possible states, i.e., motion-free and motion-affected images to

grade the intensity of motion artifacts in a CT image [21,12,26]. During training, this requires a representative labeled data set including images of all possible expected motion states and their respective motion score.

We aim to emulate the human observer with a network that identifies motion artifacts after training on clean images only. During the training process, the network learns to model the distribution of clean head CT images. Motion artifacts can then be quantified by the deviation of a sample from this distribution. In this work, we evaluate the likelihood of a given, potentially motion-affected CT image under the distribution of motion-free images represented by the trained model. Not only does this technique allow for accurate motion artifact quantification, it additionally yields enough information for gradient-based optimization of the underlying motion patterns. This allows for motion artifact compensation simply by pulling the image closer to the distribution of motion-free samples.

**Motion compensation:** Motion artifacts occur when the patient moves during the acquisition of a CT scan. This introduces a mismatch between the measured CT geometry and the one assumed during reconstruction resulting in streaks, double edges, and blurring in the final image. Most works for motion artifact reduction estimate the underlying motion pattern of a specific scan followed by a motion-compensated reconstruction. The motion estimate can be obtained from external measurements or attached markers [18,6]. Alternatively, existing approaches optimize consistency conditions on the projection data [29,2,1] or evaluate the quality of intermediate reconstructions in the image domain [14,23,21,12]. Motion compensation using diffusion models has so far only been explored for MRI [16]. The key disadvantage of consistency-based solutions is their requirement for non-truncated projection images. Consequently, this work focuses on the estimation of rigid head motion by iteratively reconstructing the data and optimizing the quality of these intermediate reconstructions.

**Out-of-distribution and anomaly detection:** Anomaly and out-of-distribution (OOD) detection are concerned with identifying samples that differ from a known distribution, mostly the training distribution [19]. These approaches often reconstruct the data through some sort of bottleneck which captures in-distribution characteristics, but fails to reconstruct OOD features [8,31,10]. Alternatively, generative networks have been employed as density estimators for the known distribution for which likelihoods can be evaluated as a proxy for the similarity of a new sample to the training data distribution [22,28,32]. Such likelihood-based methods have been proposed using normalizing flow models, energy-based models, or variational autoencoders. However, for score-based diffusion models, most existing work on OOD or anomaly detection is focused around reconstruction [27,17,7] rather than likelihood evaluation [9].

**Contributions:** In this paper, we present a motion correction method that is solely trained on a data set of clean CT images. The method thus circumvents the construction of data sets with motion artifacts that represent all possible motion states. Our contributions are (1) the grading of an image’s motion-affectedness via the exact likelihood of a diffusion model trained on motion-free images, (2) the application of a neural ordinary differential equation (ODE) solver [4] in-

side the likelihood function to calculate the gradient of the likelihood value with respect to the image, and (3) the application of this differentiable likelihood function as an objective for state-of-the-art, gradient-based motion compensation. The source code is available at [https://github.com/mareikethies/moco\\_diff\\_likelihood](https://github.com/mareikethies/moco_diff_likelihood).

## 2 Methods

### 2.1 Score-Based Diffusion Models

Score-based diffusion models approximate a data distribution by estimating gradients of the corresponding probability density [24]. Based on a diffusion process, they transform the data distribution  $p_{\text{data}}$  at time  $t = 0$  into an easily tractable prior distribution  $p_{\text{prior}}$  at time  $t = T$ . For image generation, this process can be reversed in time and samples from the prior distribution are transformed into samples from the data distribution by solving a stochastic differential equation (SDE) backward in time. Interestingly, for each such SDE, there exists a corresponding ordinary differential equation (ODE) which is deterministic and shares the same marginal distributions  $p_t(\mathbf{x})$  at any time point  $t$

$$d\mathbf{x} = \left[ \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right] dt . \quad (1)$$

This ODE is frequently referred to as probability flow ODE. Using a variance exploding variant, the drift coefficient is defined as  $\mathbf{f}(\mathbf{x}, t) = \mathbf{0}$  and the diffusion coefficient is  $g(t) = \sigma_t \sqrt{2 \log \frac{\sigma_{\max}}{\sigma_{\min}}}$  where  $\sigma_t \in [\sigma_{\min}, \sigma_{\max}]$  denotes the noise level at time point  $t$ . The gradient of the log probability density function  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$  can be approximated by a time-conditional score network  $\mathbf{s}_{\theta}$  such that  $\mathbf{s}_{\theta}(\mathbf{x}, t) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ . We train a smaller version of the originally proposed NCSN++ architecture with  $\sim 9.7 \times 10^6$  parameters,  $\sigma_{\min} = 0.01$ ,  $\sigma_{\max} = 378$  (original settings for the LSUN data set),  $10^5$  iterations, a learning rate of  $2 \times 10^{-4}$ , and a batch size of 16. Please refer to [24] for details on the training process of  $\mathbf{s}_{\theta}$ .

### 2.2 Likelihood Computation

Given a trained score network, we can generate a sample  $\mathbf{x}(0)$  from the data distribution by integrating Equation 1 backward in time starting from an initial sample  $\mathbf{x}(T)$  from the prior distribution

$$\mathbf{x}(0) = \mathbf{x}(T) + \int_T^0 \underbrace{\left[ \mathbf{f}(\mathbf{x}, t) - \frac{1}{2}g(t)^2 \mathbf{s}_{\theta}(\mathbf{x}, t) \right]}_{\tilde{\mathbf{f}}_{\theta}(\mathbf{x}, t)} dt . \quad (2)$$

Now, we can apply the continuous change of variables formula to this implicit transformation between prior and data distribution to compute the log likelihood

of  $\mathbf{x}(0)$  [24,4,15]

$$\log p(\mathbf{x}(0)) = \log p(\mathbf{x}(T)) + \int_0^T \left[ \text{tr} \left( \nabla \tilde{\mathbf{f}}_\theta(\mathbf{x}, t) \right) \right] dt . \quad (3)$$

Since the trace of the Jacobian  $\text{tr}(\nabla \tilde{\mathbf{f}}_\theta(\mathbf{x}, t))$  is computationally expensive, we follow other existing approaches [24,11] and approximate it with the Hutchinson trace estimator [13]

$$\text{tr} \left( \nabla \tilde{\mathbf{f}}_\theta(\mathbf{x}, t) \right) = \mathbb{E}_\epsilon \left[ \epsilon^T \nabla \tilde{\mathbf{f}}_\theta(\mathbf{x}, t) \epsilon \right] \approx \frac{1}{M} \sum_{m=1}^M \epsilon_m^T \nabla \tilde{\mathbf{f}}_\theta(\mathbf{x}, t) \epsilon_m , \quad (4)$$

which can be evaluated efficiently using automatic differentiation for a limited number  $M$  of random vectors  $\epsilon_m$  drawn from the Rademacher distribution. Ultimately, we can evaluate the likelihood in Equation 3 for any given data sample  $\mathbf{x}(0)$  by solving an ODE forward in time. In [24], Song et al. use an adaptive step size Runge-Kutta solver of order five from the *scipy* library and  $M = 5$  realizations of random vectors  $\epsilon_m$  to solve the ODE with high accuracy.

### 2.3 Neural ODEs and Likelihood Target Function

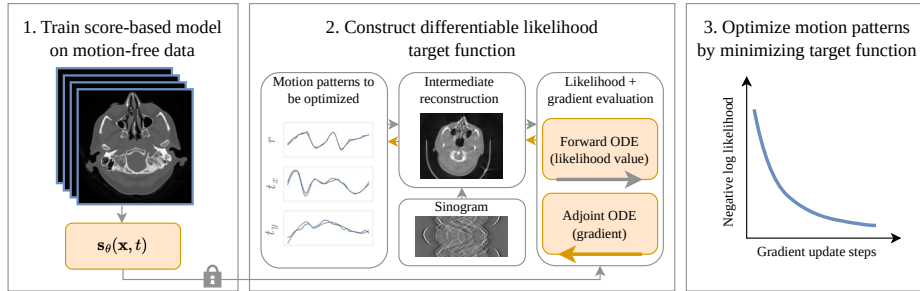
In this work, we use Equation 3 as the objective function in a gradient-based optimization problem for CT motion compensation. This entails two implications: (1) We can not simply utilize any ODE solver, but we require differentiability through the solver to obtain the gradient of the log likelihood  $\log p(\mathbf{x}(0))$  with respect to the image  $\mathbf{x}(0)$ , and (2) we do not necessarily need to solve Equation 3 to the highest possible precision as long as the loss landscape remains well-behaved.

To address these requirements, we substitute the conventional ODE solver with a neural ODE solver<sup>3</sup> to ensure differentiability. Instead of differentiating through the internals of the solver by tracking the computations of the forward pass, it solves the so-called adjoint ODE for gradient computation [4]. This ensures differentiable behavior at low memory cost since intermediate results do not need to be stored. To reduce the computation time, we configure a fixed step size Runge-Kutta solver of order four with ten steps for the forward ODE and 20 steps for the gradient calculation via the adjoint ODE. Moreover, for each evaluation of Equation 3, we sample one realization of  $\epsilon_m$  and keep it constant for the gradient computation. Repeated evaluations of the target function use different random vectors  $\epsilon_m$ .

### 2.4 Motion Compensation

Figure 1 summarizes the proposed algorithm using the differentiable likelihood function as the target function for motion compensation. First, a score network

<sup>3</sup> <https://github.com/rtqichen/torchdiffeq>



**Fig. 1.** Overview of the proposed pipeline. After training on motion-free images, the score model  $\mathbf{s}_\theta(\mathbf{x}, t)$  is used inside the likelihood target function. For each step of the gradient descent optimizer, we perform an intermediate reconstruction and evaluate the gradient of the likelihood function to pull the image closer to the distribution of clean images seen during training.

is fitted to the distribution of motion-free images which is subsequently applied inside the likelihood target function in Equation 3. For motion compensation, we estimate a set of Akima splines describing smooth and non-overshooting motion trajectories. Each spline parameterizes the values for one of the three rigid motion parameters (translations  $t_x$  and  $t_y$  and rotation  $r$ ) throughout the scan with 30 evenly distributed nodes yielding 90 free parameters  $\gamma^{(n)}$  in total for the optimization problem at step  $n$ . The motion parameters are incorporated into the projection matrices defining the fan-beam geometry by matrix multiplication. In each iteration of the optimization algorithm, we (1) apply the current state of motion parameters to the fan-beam geometry, (2) reconstruct the sinogram based on this trajectory to obtain an intermediate, possibly motion-affected image  $\mathbf{x}^{(n)}(0)$ , and (3) feed this image into the previously introduced likelihood function to grade its motion-affectedness in terms of a negative log likelihood value. Next, the solution to the adjoint ODE yields the gradient of the likelihood value with respect to the intermediate reconstructed image  $\frac{d \log p(\mathbf{x}^{(n)}(0))}{d\mathbf{x}^{(n)}(0)}$  which is subsequently propagated into the free motion parameters using the Jacobian of the fan-beam reconstruction operator  $\frac{d\mathbf{x}^{(n)}(0)}{d\gamma^{(n)}}$  [25,26]

$$\frac{d \log p(\mathbf{x}^{(n)}(0))}{d\gamma^{(n)}} = \frac{d \log p(\mathbf{x}^{(n)}(0))}{d\mathbf{x}^{(n)}(0)} \cdot \frac{d\mathbf{x}^{(n)}(0)}{d\gamma^{(n)}}. \quad (5)$$

Given the target function's gradient with respect to the free parameters, the spline-based motion patterns are optimized using basic gradient descent

$$\gamma^{(n+1)} = \gamma^{(n)} + r^{(n)} \cdot \frac{d \log p(\mathbf{x}^{(n)}(0))}{d\gamma^{(n)}} \quad (6)$$

with an initial estimate  $\gamma^{(0)} = \mathbf{0}$  and an exponentially decaying step size  $r^{(n)} = r_0 \cdot q^n$ . 40 iterations of Equation 6 are executed with an initial step size of  $r_0 = 100$  and a decay factor of  $q = 0.97$ .

### 3 Experiments and Results

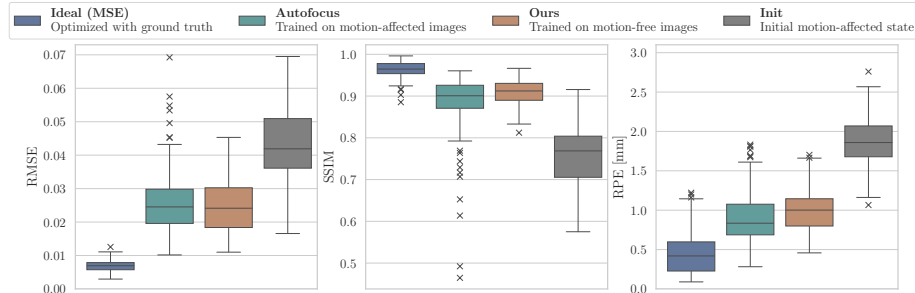
Our experiments are performed retrospectively on publicly available head CT scans from the CQ500 data set published under CC BY-NC-SA license [5]. All terms of use in the end-user license agreement were followed strictly. The score network is trained on slices from 200 subjects and evaluated on slices from 40 subjects. The motion compensation experiments are performed on slices from another 40 disjunct subjects. For the score network, we use a reduced-parameter version of the NCSN++ architecture proposed in [24]. Exemplary generated head CT samples from the trained model can be found in the supplementary material.

Motion compensation is performed on individual slices. A sinogram is simulated for each slice by forward projecting it onto a trajectory with 360 projections on a full circle, a source-to-isocenter distance of 785 mm, a source-to-detector distance of 1200 mm, and a detector with 700 elements at spacing 0.64 mm. Reconstructions are computed on a grid of  $256 \times 256$  pixels with an isotropic spacing of 1 mm. The initial motion-affected state is obtained by perturbing the ground-truth reconstruction trajectory with a spline-based motion pattern with ten nodes and a maximal amplitude of 5 mm for translations  $t_x$  and  $t_y$  as well as  $5^\circ$  for rotation  $r$ . The likelihood-based target function is compared to two alternative objectives. Importantly, both reference methods require substantially more knowledge about the problem than our proposed method. First, as an upper performance bound, motion parameters are optimized by computing the mean-squared error (MSE) to the ground-truth, motion-free reconstruction. This represents an ideal target function, but requires knowledge of the motion-free image at optimization time, making it impractical for almost all applications. Second, following existing approaches [25,21,12], we train a network on a paired data set of motion-affected and motion-free images to predict the structural similarity index measure (SSIM) given only the motion-affected image. This trained network is referred to as autofocus target function in the following. Notably, in contrast to the proposed target function, the autofocus network has seen examples of motion-affected images during training.

**Table 1.** Average quantitative values for all investigated metrics comparing the proposed method to the autofocus target function and the initial motion-affected reconstruction. The ideal MSE target function is excluded since it requires knowledge of the motion-free ground truth at optimization time and is therefore impractical. The best value is highlighted in bold font.

	RMSE [ $\times 0.1$ ] ( $\downarrow$ )	SSIM ( $\uparrow$ )	RPE ( $\downarrow$ )	MAE $t_x$ ( $\downarrow$ )	MAE $t_y$ ( $\downarrow$ )	MAE $r$ ( $\downarrow$ )
Init	0.45	0.74	1.98	1.01	1.02	0.98
Autofocus	0.26	0.89	<b>0.94</b>	0.75	0.77	<b>0.44</b>
Ours	<b>0.25</b>	<b>0.91</b>	0.99	<b>0.71</b>	<b>0.72</b>	0.51

Figure 2 illustrates the motion compensation performance of the investigated target functions. Overall, all methods improve upon the initial motion-affected

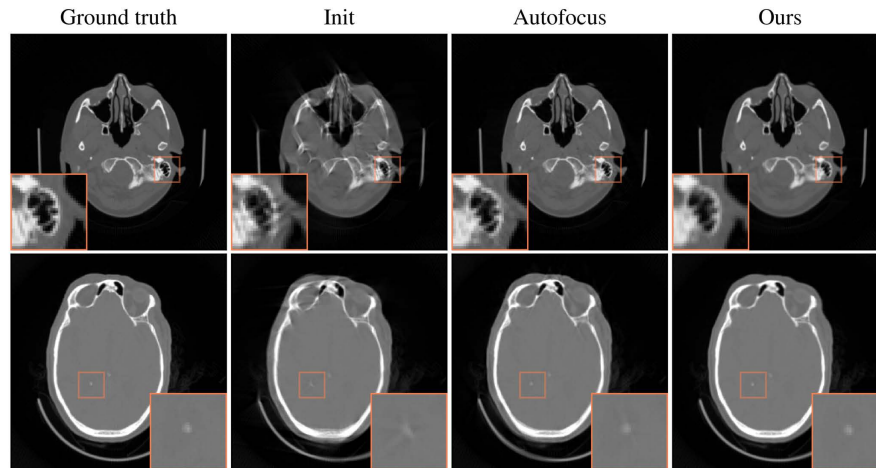


**Fig. 2.** Comparison of target functions using root mean squared error (RMSE) ( $\downarrow$ ) and SSIM ( $\uparrow$ ) of the motion-compensated images as well as reprojection error (RPE) ( $\downarrow$ ). The proposed method achieves similar results as the autofocus method despite never having seen any motion-affected images during training. The MSE target function can be considered an upper performance bound since it requires knowledge of the motion-free image at optimization time which is unrealistic in a clinical workflow.

state. The MSE-based optimization with known ground truth performs best. This is to be expected and can be regarded as the upper performance bound of the motion compensation pipeline given an ideal target function. The autofocus target function and the likelihood-based objective proposed in this work perform similarly concerning the RMSE and SSIM of the motion-compensated images. The autofocus method results in a slightly lower RPE, but higher variance and more negative outliers. These findings are confirmed in Table 1. Additionally, we compute the mean absolute error (MAE) for each of the three motion parameters ( $t_x$ ,  $t_y$ , and  $r$ ). While the likelihood objective yields a higher MAE for rotation, it recovers translations slightly better than the autofocus approach. Figure 3 shows two example slices before and after motion compensation with the autofocus and likelihood target function. Particularly in the enlarged regions of interest, a clear improvement over the initial motion-affected state can be observed with fewer streaks and less blur. The compensated image from the likelihood objective appears slightly smoother than the one from the autofocus objective, but both successfully restore even fine details such as the bright spot in the bottom example.

## 4 Discussion

The proposed method performs on par with the autofocus approach despite the fact that the autofocus network was trained on motion-affected examples of different motion patterns and motion amplitudes. In contrast, the proposed motion compensation method only needs motion-free images. Practically, this is an enormous difference since multiple data sets of clean CT images are publicly available. On the other hand, creating a paired data set of motion-free and motion-affected data as needed by the autofocus-type approaches is not



**Fig. 3.** Two example slices before and after motion compensation with autofocus and likelihood objective. A  $\times 3$  zoom of a region of interest is shown in the orange frame.

straightforward. The simulation of representative motion patterns that are applied to the projection geometry for motion perturbation is complex and usually based on more or less realistic assumptions about the actual motion of patients in a scanner. As a result, specific motion patterns occurring in practice might not be covered. The likelihood-based approach is free of such assumptions. We acknowledge that evaluating the proposed target function and its gradient is considerably slower than the other objectives because it needs to solve two ODEs per gradient update step which both depend on the score network. Hence, the forward and backward pass of the score network are evaluated multiple times for a single update of the motion parameters. This property is inherited from the diffusion process in score-based modeling which also makes sample generation slower compared to other generative models. However, speeding up the sampling is an active area of research and we hypothesize that such approaches could improve the speed of likelihood computation as well [30]. Additionally, we would like to comment on the findings of existing papers which generally question the usability of likelihoods for OOD or anomaly detection since they seem to be systematically biased by the complexity of the data [20,3]. Whereas our proposed target function is inspired by the ideas of likelihood-based OOD detection, we only perform a per-sample optimization for motion compensation. Consequently, we do not require a total separation of motion-free and motion-affected image distributions as long as the motion-free state of any given individual sample is graded better than a motion-affected state of the same sample.



## 5 Conclusion

In this work, we extend the exact likelihood computation of score-based models, making it applicable as a target function in gradient-based optimization. Applied to CT motion compensation, it eliminates the need to simulate examples of motion-affected images for training. Since we only model the distribution of clean images, the same target function could be envisioned for other image restoration tasks without changes posing an exciting direction for future research.

**Acknowledgments.** The research leading to these results has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (ERC Grant No. 810316). The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center of the Friedrich-Alexander-Universität Erlangen-Nürnberg. The hardware is funded by the German Research Foundation.

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

## References

1. Aichert, A., Berger, M., Wang, J., Maass, N., Doerfler, A., Hornegger, J., Maier, A.K.: Epipolar consistency in transmission imaging. *IEEE transactions on medical imaging* **34**(11), 2205–2219 (2015)
2. Berger, M., Xia, Y., Aichinger, W., Mentl, K., Unberath, M., Aichert, A., Riess, C., Hornegger, J., Fahrig, R., Maier, A.: Motion compensation for cone-beam CT using Fourier consistency conditions. *Physics in Medicine & Biology* **62**(17), 7181 (2017)
3. Caterini, A.L., Loaiza-Ganem, G.: Entropic issues in likelihood-based ood detection. In: *I (Still) Can't Believe It's Not Better! Workshop at NeurIPS 2021*. pp. 21–26. PMLR (2022)
4. Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural Ordinary Differential Equations. *Advances in neural information processing systems* **31** (2018)
5. Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N.G., Venugopal, V.K., Mahajan, V., Rao, P., Warier, P.: Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet* **392**(10162), 2388–2396 (2018), <http://headctstudy.que.ai/dataset>
6. Choi, J.H., Fahrig, R., Keil, A., Besier, T.F., Pal, S., McWalter, E.J., Beaupré, G.S., Maier, A.: Fiducial marker-based correction for involuntary motion in weight-bearing C-arm CT scanning of knees. Part I. Numerical model-based optimization. *Medical physics* **40**(9), 091905 (2013)
7. Choi, S., Lee, H., Lee, H., Lee, M.: Projection Regret: Reducing Background Bias for Novelty Detection via Diffusion Models. *Advances in Neural Information Processing Systems* **36** (2024)
8. Denouden, T., Salay, R., Czarnecki, K., Abdelzad, V., Phan, B., Vernekar, S.: Improving Reconstruction Autoencoder Out-of-distribution Detection with Mahalanobis Distance. *arXiv preprint arXiv:1812.02765* (2018)
9. Goodier, J., Campbell, N.D.: Likelihood-based Out-of-Distribution Detection with Denoising Diffusion Probabilistic Models. *arXiv preprint arXiv:2310.17432* (2023)
10. Graham, M.S., Pinaya, W.H.L., Wright, P., Tudosiu, P.D., Mah, Y.H., Teo, J.T., Jäger, H.R., Werring, D., et al.: Unsupervised 3D Out-of-Distribution Detection with Latent Diffusion Models. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 446–456. Springer (2023)
11. Grathwohl, W., Chen, R.T., Bettencourt, J., Sutskever, I., Duvenaud, D.: FFJORD: Free-Form Continuous Dynamics for Scalable Reversible Generative Models. In: *International Conference on Learning Representations* (2018)
12. Huang, H., Siewerdsen, J.H., Zbijewski, W., Weiss, C.R., Unberath, M., Ehtiyati, T., Sisniega, A.: Reference-free learning-based similarity metric for motion compensation in cone-beam CT. *Physics in Medicine & Biology* **67**(12), 125020 (2022)
13. Hutchinson, M.F.: A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation* **18**(3), 1059–1076 (1989)
14. Kingston, A., Sakellariou, A., Varslot, T., Myers, G., Sheppard, A.: Reliable automatic alignment of tomographic projection data by passive auto-focus. *Medical physics* **38**(9), 4934–4945 (2011)
15. Köthe, U.: A review of change of variable formulas for generative modeling. *arXiv preprint arXiv:2308.02652* (2023)
16. Levac, B., Kumar, S., Jalal, A., Tamir, J.I.: Accelerated motion correction with deep generative diffusion models. *Magnetic Resonance in Medicine* (2024)

17. Linmans, J., Raya, G., van der Laak, J., Litjens, G.: Diffusion models for out-of-distribution detection in digital pathology. *Medical Image Analysis* **93**, 103088 (2024)
18. Maier, J., Nitschke, M., Choi, J.H., Gold, G., Fahrig, R., Eskofier, B.M., Maier, A.: Inertial measurements for motion compensation in weight-bearing cone-beam CT of the knee. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 14–23. Springer (2020)
19. Müller, J.P., Baugh, M., Tan, J., Dombrowski, M., Kainz, B.: Confidence-Aware and Self-supervised Image Anomaly Localisation. In: *International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*. pp. 177–187. Springer (2023)
20. Nalisnick, E., Matsukawa, A., Teh, Y.W., Gorur, D., Lakshminarayanan, B.: Do deep generative models know what they don't know? In: *International Conference on Learning Representations* (2019)
21. Preuhs, A., Manhart, M., Roser, P., Hoppe, E., Huang, Y., Psychogios, M., Kowarschik, M., Maier, A.: Appearance learning for image-based motion estimation in tomography. *IEEE Transactions on Medical Imaging* **39**(11), 3667–3678 (2020)
22. Ren, J., Liu, P.J., Fertig, E., Snoek, J., Poplin, R., Deprieto, M., Dillon, J., Lakshminarayanan, B.: Likelihood Ratios for Out-of-Distribution Detection. *Advances in neural information processing systems* **32** (2019)
23. Sisniega, A., Stayman, J.W., Yorkston, J., Siewerdsen, J., Zbijewski, W.: Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion. *Physics in Medicine & Biology* **62**(9), 3712 (2017)
24. 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)
25. Thies, M., Wagner, F., Maul, N., Folle, L., Meier, M., Rohleder, M., Schneider, L.S., Pfaff, L., Gu, M., Utz, J., et al.: Gradient-based geometry learning for fan-beam CT reconstruction. *Physics in Medicine & Biology* **68**(20), 205004 (2023)
26. Thies, M., Wagner, F., Maul, N., Yu, H., Meier, M., Schneider, L.S., Gu, M., Mei, S., Folle, L., Maier, A.: A gradient-based approach to fast and accurate head motion compensation in cone-beam CT. *arXiv preprint arXiv:2401.09283* (2024)
27. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion Models for Medical Anomaly Detection. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 35–45. Springer (2022)
28. Xiao, Z., Yan, Q., Amit, Y.: Likelihood Regret: An Out-of-Distribution Detection Score For Variational Auto-encoder. *Advances in neural information processing systems* **33**, 20685–20696 (2020)
29. Yu, H., Wang, G.: Data consistency based rigid motion artifact reduction in fan-beam CT. *IEEE transactions on medical imaging* **26**(2), 249–260 (2007)
30. Zheng, H., Nie, W., Vahdat, A., Azizzadenesheli, K., Anandkumar, A.: Fast Sampling of Diffusion Models via Operator Learning. In: *NeurIPS 2022 Workshop on Score-Based Methods* (2022)
31. Zhou, Y.: Rethinking Reconstruction Autoencoder-Based Out-of-Distribution Detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7379–7387 (2022)
32. Zisselman, E., Tamar, A.: Deep Residual Flow for Out of Distribution Detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 13994–14003 (2020)