Upper Error Bounds for Score-Based Inverse Problem Solving in Imaging

Irina Dobrianski [1] **Dominik Narnhofer** [2] **Thomas Pock** [3]

Abstract

Diffusion models have gained tremendous popularity for generating diverse and high quality images. We adapt diffusion models for solving inverse problems in imaging. Moreover, we quantify the uncertainties in the reconstructed images by deriving pixel-wise upper error bounds dependent on the determined variance without relying on the ground-truth. Especially in high-stake applications such as healthcare, well-calibrated uncertainties are vital for reliable decision making. For example in magnetic resonance imaging, undersampling the k-space plays a crucial role in clinical applications and it is highly important to be aware of the uncertainties as the diagnosis and further treatments depend on the reconstructed images. In this work, we focus on the score-based generative models through stochastic differential equations and show that an unconditional diffusion model trained on a specific dataset (BSDS and fastMRI) can be utilized for solving various inverse problems, e.g. denoising, inpainting and zero-filling and the uncertainty quantification yields a strong correlation between the squared error and the variance.

1. Introduction

Diffusion models are well-known for their ability to generate high-quality samples and capture large diversity. However, when handling inverse problems, it becomes crucial to acquire knowledge about both the extent and the precise location of uncertainties in the reconstructed image. Ill-posed inverse problems arise in various applications in the field of computer vision, e.g. denoising, inpainting, super-resolution, colorization, JPEG restoration, Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) to name just a few of them (Yang et al., 2022). Especially in the medical field, the diagnosis and also further treatments depend on the quality of the reconstructed images. There are several approaches that compute the variance as a measure for the uncertainty (Luo et al., 2023; Wolleb et al., 2021; Xie & Li, 2022). However, they do not relate the variance to the error between the reconstructed and the original image. This shortfall is handled in Narnhofer et al. (2022) where our approach is strongly based on. We apply this promising method in combination with the emerging diffusion models which are known for their good mode coverage and high-quality samples and will demonstrate that we can derive upper error bound guarantees depending on the variance.

Diffusion models operate in two stages. In the first stage, the forward process, input data gets progressively corrupted with noise until pure Gaussian noise with zero mean and unit variance is reached whereby the model is trained to predict the added noise. In the second stage, the reverse process, the corrupted data is denoised again step-by-step. Diffusion models can be subdivided into Denoising Diffusion Probabilistic Models (DDPMs), Score-Based Generative Models (SBGMs) and SBGMs through Stochastic Differential Equations (SDEs) (Song et al., 2021). The SBGMs through SDEs predict the score which is the gradient of the logarithmic probability density. We focus on the SBGMs through Variance Preserving (VP) SDEs which can be related to the DDPMs.

To verify the approach we consider three inverse problems: denoising, inpainting and zero-filling. In the denoising task, the goal is to remove the noise which is particularly important as noise is part of every measurement process also when capturing images (Szelsiki, 2022). The objective of the inpainting problem is to reconstruct missing information. For MRI, undersampling the k-space and thus violating the Nyquist criterion leads to aliasing artifacts in the reconstructed image. However, it reduces the acquisition time and further increases the patient throughput which also helps to save costs (Peng et al., 2022). Moreover, reducing the acquisition time increases the patient's comfort and also reduces motion artifacts as the patient's tendency to move increases with time. We apply Bayes' theorem to combine diffusion models with inverse problems to solve them. The

[1]Institute of Theoretical Computer Science, Graz University of Technology, Austria [2]Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) Zurich [3]Institute for Computer Graphic and Vision, Graz University of Technology, Austria. Correspondence to: Irina Dobrianski <irina.dobrianski@tugraz.at>, Dominik Narnhofer <dominik.narnhofer@geod.baug.ethz.ch>, Thomas Pock <thomas.pock@tugraz.at>.

Accepted by the Structured Probabilistic Inference & Generative Modeling workshop of ICML 2024, Vienna, Austria. Copyright 2024 by the author(s).

posterior distribution of the clean image $x \in \mathbb{R}^{M \times N}$ of size $M \times N$ given the corrupted image $z \in \mathbb{R}^{M \times N}$ is according to Bayes' theorem $p(x|z) \propto p(x)p(z|x)$ where $p(x)$ is the prior and $p(z|x)$ the likelihood term. The trained diffusion model serves as prior and the inverse problem defines the likelihood term itself. Usually linear inverse problems are defined as $z = Ax + \epsilon$ where A is the forward operator and $\epsilon \sim \mathcal{N}(0, I)$. For solving the SDEs we apply the Euler-Maruyama method and the predictor method from Song et al. (2021). Due to the stochastic property of the diffusion models and sampling methods we can compute the Minimum Mean Squared Error (MMSE) and the variance from several samples.

To generate several samples we either use the single or the multiple trajectories method. For the single trajectory method we compute the metrics from the samples of the last thousands steps of the reverse diffusion process whereby we keep the variance of the added noise constant for these steps. For the multiple trajectories method we split the reverse process after a few hundred steps and then perform the reverse process for each sample separately. With the computed variance and the squared error between the MMSE and the original image we obtain pixel-wise error bounds according to the method proposed in Narnhofer et al. (2022) and can show that there is a relation between the variance and the squared error. We apply the error quantification on the Berkeley Segmentation DataSet (BSDS) (Martin et al., 2001) and on the fastMRI (Zbontar et al., 2018) dataset. For the three inverse problems, denoising, inpainting and zero-filling, and both datasets, BSDS and fastMRI, we obtain higher error bounds for higher variance values and vice versa. Hence, if a specific region in an image is important but contains high variance values the outcome and decisions may be reconsidered due to the higher predicted error quantile.

2. Related Work

Previous research has been carried out in quantifying the uncertainty using the variance of reconstructed images with diffusion models. However, the variance is rarely compared with the error whereas we provide upper error bound guarantees based on the method proposed by Narnhofer et al. (2022) combined with the popular diffusion models. Other recent studies neglect the correlation between the variance and the error. Ramzi et al. (2020) display solely that the reconstructed samples vary especially in certain areas without computing any further metrics. Xie & Li (2022) compute the pixel-wise variance of several samples of the posterior distribution and Jalal et al. (2021) compute the standard deviation and compare it to the error image. A further description how the error and the variance relate is not stated. A short statement about uncertainty can be found in Chung & Ye (2021). They obtain higher variance values for higher acceleration factors and state that with larger acceleration factors also the uncertainty is higher. A more detailed explanation is declared in Luo et al. (2023) where the variance as well as the 95% interval is computed and it is figured out that the variance is high at edges and regions where aliasing occurs and related to the missing high-frequency information in the undersampled k-space but it is not compared or related to the error.

3. Theory

Inverse problems are part of numerous applications and areas in science like geophysics, chemistry, astronomy, medical imaging, optics and acoustic (Idier, 2008; Kern, 2016). In this work the focus is on the relevant aspects for imaging. Mathematically, a forward process of a linear inverse problem can generally be described by $z = Ax + \epsilon$ where $z \in \mathcal{Z}$ is the measurement data, $A : \mathcal{X} \to \mathcal{Z}$ the forward operator and $x \in \mathcal{X}$ the ground-truth signal. Additionally, every measurement acquisition is corrupted by noise ϵ. The forward operator A often cannot be inverted or is ill-conditioned and therefore the inverse problem is called ill-posed according to the definition of Hadamard.

In the denoising task we face the problem that the ground-truth image is corrupted by noise and the goal is to recover a clean image from the noisy observation. We assume zero mean additive Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with variance σ^2. The inpainting problem is characterized by restoring missing parts of an image. The forward operator A corresponds to a binary mask which is zero for unknown pixels and one otherwise. The zero-filling task is only relevant for MR data. As MRI suffers from a high acquisition time leading to a lower patient throughput, motion artifacts and overall high costs, lines in the k-space can be skipped and filled with zeros. However, undersampling the k-space causes aliasing artifacts as the Nyquist-Shannon theorem is violated (Zbontar et al., 2018). In this task we deal with a binary mask in the Fourier space rather than in the image space.

Diffusion models are generative models and applied in various fields, e.g. computer vision, natural language processing, multi-modal generation and temporal data modeling (Yang et al., 2022). The working flow of diffusion models can be separated in a forward and a reverse process whereby in the forward process noise is added to the data step-by-step and in the reverse process the goal is to reconstruct samples from the learned distribution by gradually denoising them (Ho et al., 2020; Song & Ermon, 2019; 2020; Song et al., 2021). In this work the focus is on the SBGMs through SDEs from Song et al. (2021). They use a diffusion process to transform data to noise which can be described by an SDE in the general form of $\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w$ where $f(x,t)$ is the drift coefficient, $g(t)$ the diffusion co-

efficient which is modeled as a scalar independent of x and w describes the Wiener process or Brownian motion. The time t is continuous between $[0, T]$ and the process starts with $x(0) \sim p_0$ where p_0 is the data distribution and ends with $x(T) \sim p_T$ where p_T is the prior distribution which is usually a Gaussian distribution with a certain mean and variance (Song et al., 2021). The forward SDE can be reversed which yields the reverse-time SDE $\mathrm{d}x = f(x,t)\mathrm{d}t - g^2(t)\nabla \log p_t(x)\mathrm{d}w + g(t)\mathrm{d}w$ where t starts with T and ends with zero and $\mathrm{d}t$ is negative (Song et al., 2021). Furthermore, $p_t(x)$ is the probability density of $x(t)$ where $x(t) \sim p_{0t}(x(t)|x(0))$ and $p_{0t}(x(t)|x(0))$ is the transition kernel which is usually Gaussian. The score function $s_\theta(x,t) \approx \nabla \log p_t(x)$ is obtained by training a score-based model whereby Song et al. (2021) define the objective of the score-matching training with the weighting function λ as

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_t \left[\lambda(t)\mathbb{E}_{x(0)}\mathbb{E}_{x(t)|x(0)} \left[\|s_\theta(x(t),t) - \nabla \log p_{0t}(x(t)|x(0))\|_2^2\right]\right]. \tag{1}$$

In this work we use the Variance Preserving SDEs which are the continuous version of the DDPMs from Ho et al. (2020). Song et al. (2021) derive the perturbation kernel for the forward process as

$$p_{0t}(x(t)|x(0)) = \mathcal{N}\Big(x(t); e^{-\frac{1}{4}t^2(\beta_{\max}-\beta_{\min})-\frac{1}{2}t\beta_{\min}}\, x(0), I - Ie^{-\frac{1}{2}t^2(\beta_{\max}-\beta_{\min})-t\beta_{\min}}\Big) \tag{2}$$

with $\{\beta_i\}_{i=1}^N$ and $\beta_i = \frac{\beta_{\min}}{N} + \frac{i-1}{N(N-1)}(\beta_{\max} - \beta_{\min})$ and for $N \to \infty$ the variance is $\beta(t) = \beta_{\min} + t(\beta_{\max} - \beta_{\min})$ and $t \in [0,1]$. The advantage of the SBGMs compared to the DDPMs is that a general SDE solver can be used in the reverse process. For solving inverse problems with the BSDS dataset we apply the Euler-Maruyama method

$$x_{t-1} = x_t + \frac{1}{2}\beta_t\left(x_t + 2s_\theta\left(x_t, t\right)\right)\Delta t + \sqrt{\beta_t \Delta t}\,\epsilon - \lambda \nabla_{x_t}\mathcal{D}(x_t, x_T) \tag{3}$$

and add the data term $\mathcal{D}(x,z) = \frac{1}{2}||Ax - z||^2$ to specify the inverse problem. For the fastMRI dataset we obtained improved results with the predictor sampler from Song et al. (2021).

Error Quantification. Inverse problems and diffusion models can be combined through Bayes' theorem which arises for an observation z and the ground-truth x to $p(x|z) = \frac{p(z|x)p(x)}{p(z)}$ where $p(x|z)$ is the posterior, $p(z|x)$ the likelihood and $p(x)$ the prior. We use diffusion models as prior and the likelihood term is defined by the inverse problem itself. The denominator is a scaling factor and does not depend on x.

The theory for the uncertainty quantification is based on the method proposed by Narnhofer et al. (2022). The aim is to get error bounds with coverage guarantees depending on the posterior variance. We do not list all the evidences here but we adhere to the notation in Narnhofer et al. (2022) for better comparison. We assume an observation Z which is an estimate for the ground-truth data X with instances z and x respectively. Furthermore, we assume a prediction function $\hat{x}(z)$ which can take different forms. In this approach we use the mean of the posterior $\hat{x}(z) \approx \mathbb{E}[X|Z = z]$ and $\hat{x}(Z)$ is the prediction for X. Also the estimator function of the posterior variance defined by $\hat{t}(z) \approx \mathrm{Var}[X|Z = z]$ is available, where $\hat{t}(Z) = \hat{T}$. The range of $\hat{T}$ is split into disjoint bins $\tau_k = [t_k, t_{k+1})$ for $k = 1, 2, 3, ...$ otherwise it cannot be used in practice as there are not infinitely many values available for a finite amount of data as stated by Narnhofer et al. (2022).

In general, the Cumulative Distribution Function (CDF) is defined by $F(y) = \mathrm{P}[Y \leq y]$ and the quantile function is the inverse CDF $F^{-1}(q) = y_q = \inf\{y \in \mathbb{R} : F(y) \geq q\}$ with $q \in (0,1)$ whereas the empirical distribution function is defined by$\hat{F}(y) = \frac{1}{n}\sum_{i=1}^n 1_{Y_i \leq y}$ and the empirical q-quantile by $\hat{Y}_q = \inf\{y \in \mathbb{R} : \hat{F}(y) \geq q\}$. The quantile can be computed by $\hat{Y}_q = Y_{(\lceil nq \rceil)}$ with $Y_{(m)}$ denoting the m-th smallest value of the samples. If we have a random sample size $N+1$ whereby the last sample is the new one then the quantile arises to $\mathrm{P}[Y_{N+1} \leq \hat{Y}_q\, |N = n] \geq q$ and can be computed with $\hat{Y}_q = Y_{(\lceil (N+1)q \rceil)}$. To guarantee the coverage for new samples the quantile is corrected by $1 + \frac{1}{N}$ and results in $\mathrm{P}[Y \leq \hat{Y}_{(1+\frac{1}{N})q}|N = n] \geq q$ which can be computed by $\hat{Y}_{(1+\frac{1}{N})q} = Y_{(\lceil (N+1)q \rceil)}$. We have i.i.d. samples $(X_i, Z_i)_{i=1}^m$ and the squared error S defined by $S = s(X,Z) = (X - \hat{x}(Z))^2$. With the basics above and the exact derivation in Narnhofer et al. (2022) it can be shown that the goal is to compute predicting quantiles $\hat{S}_q = \hat{s}_q((X_i, Z_i)_{i=1}^m, Z)$ such that $\mathrm{P}[S \leq \hat{S}_q] \geq q$ with the quantile $q \in (0,1)$. Considering the discretization of the range of $\hat{T}$ the estimated conditional error quantile is defined as $\hat{S}_q = \left(1 + \frac{1}{N_{\tau_k}}\right) q$ with $N_{\tau_k} = |\{S_i | \hat{T}_i \in \tau_k\}|$ and if $(1 + N_{\tau_k})q \geq N_{\tau_k}$ it is the essential supremum of S. For evaluation purposes we check if the true errors are smaller than the predicted error quantiles. This means that the mean of the binary decisions should correspond to the selected quantile which is measured by the coverage introduced by Narnhofer et al. (2022).

4. Experiments

We conducted a series of experiments and will list the most important findings below and make the results comparable to the method proposed in Narnhofer et al. (2022). The experimental settings can be found in Appendix A.

We evaluate the reconstructed images by computing the Peak Signal-to-Noise Ratio (PSNR), the mutual information between the variance and the error of the training images and the coverage of the test images for the given quantile as stated in Narnhofer et al. (2022). Our results (see Table 1 and 2 in Appendix A) cannot outperform those of the reference method in Narnhofer et al. (2022) but are very close to it. We have to be careful when comparing these two approaches as already the PSNR values of the corrupted images are not consistent with each other and for the denoising task the used noise levels differ in 0.5%. Furthermore, they used only 68 test images instead of 100 for the BSDS dataset. For the fastMRI dataset, we do not know if we evaluate the methods on the same middle slices in each volume as the exact slice is not explicitly stated. Another unknown aspect is whether the zero-filling procedure follows a random approach, as we opted for, or an equidistant one. A notable distinction is that our model is also trained on the PDFS images. Even though the evaluation process does not take the PDFS images into account we can still solve inverse problems with this kind of images. Thus, our model demonstrates higher generalization capabilities. This may also explain the marginally lower values as our model is not specially trained on the PD images. Narnhofer et al. (2022) use 50k reverse steps which is approximately six times more than we run and we do not need any burn in phases.

Figure 1 shows the upper error bound guarantees for the MR undersampling task with $R = 4$. The figures for the other inverse problems as well as the corrupted and reconstructed images are in Appendix A. Considering all three inverse problems and the used datasets, we observe that a higher corruption leads to a more difficult reconstruction. Also the PSNR values of the reconstructed images decrease, as expected, with greater distortion of the original image. Comparing the single and the multiple trajectories method, the former produces higher variance values as there is always some small noise added in the sampling procedure. For the multiple trajectories method there is no noise added in the last step. This approach produces a higher linear dependence between the error and the variance. The oscillations at the minimum and maximum values are probably due to fewer samples in these regions. The disadvantage of the multiple trajectories method is the slower sampling process. Therefore, we generated fewer samples for computing the MMSE and the variance. On the other hand, the downside of the single trajectory method is that this procedure tends to produce smeary outcomes.

Denoising. The error images are noisier than the predicted 0.9 quantile (see Figure 3). However, both show the structures and borders of the objects. Furthermore, the estimated error quantile increases with higher variance values.

Inpainting. The mutual information is the highest for the inpainting task as there are mainly the mask structures relevant. The error and the predicted quantile images show the structures of the masks which have higher values in heterogeneous regions of the image.

Zero-Filling. In practice, acquisition time is reduced by skipping lines of the k-space but this leads to aliasing artifacts when reconstructing the images by applying the inverse Fourier transform to the k-space. This high-frequency information is masked by the zero-filling task, therefore we obtain higher uncertainty in fine bone structures in the reconstructed images. Furthermore, we show that the uncertainty values increase with larger acceleration factors as there is even more high-frequency information missing.

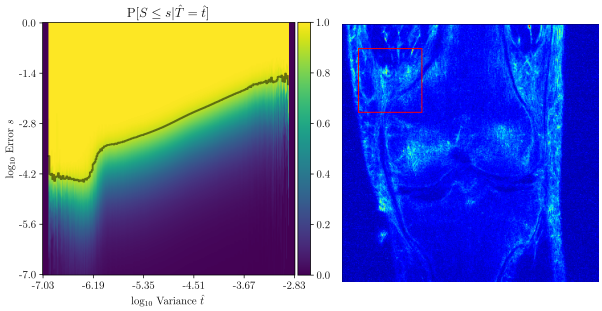

Figure 1. Left: The cumulative conditional distribution $P[S \leq s|\hat{T} = \hat{t}]$ of the error s and the variance $\hat{t}$ as well as the 0.9 quantile of the zero-filling task for $R = 4$. Right: The 0.9 quantile image with the range 4e-2.

5. Conclusion

We applied diffusion models as prior in a variational Bayesian setting for solving inverse problems. The specific inverse problem defined the likelihood term itself. The advantage of this approach is that we only need to train one model for a specific dataset and can solve various inverse problems. With the stochastic sampling methods we compute the MMSE and the variance which are further used to obtain upper error bounds. Being aware of the uncertainty when reconstructing images is especially important in the medical sector. The uncertainty may play a crucial role for accurate diagnosis and effective therapies. However, the advantage of gaining knowledge about the uncertainty is not yet examined in daily clinical practice. Future work may empirically evaluate the benefits of uncertainty quantification that doctors perceive. Furthermore, more sophisticated model architectures can be incorporated and the reconstruction processes further enhanced and sped up. Refining the reconstruction process can lead to improved PSNR values that can keep up with the values proposed in Narnhofer et al. (2022). Moreover, with better reconstruction results the uncertainty estimation will gain higher precision and even a higher correlation to the error may be achieved.

References

Chung, H. and Ye, J. C. Score-based diffusion models for accelerated MRI. *Medical image analysis*, 80:102479, 2021.

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models. In *Proceedings of the International Conference on Neural Information Processing Systems*, 2020.

Huang, C.-W. sdeflow-light. https://github.com/CW-Huang/sdeflow-light, 2021. Online; accessed February 05, 2023.

Idier, J. *Bayesian Approach to Inverse Problems*. ISTE, first edition, 2008.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A. G., and Tamir, J. Robust Compressed Sensing MRI with Deep Generative Priors. In *Proceedings of the International Conference on Neural Information Processing Systems*, 2021.

Kern, M. *Numerical Methods for Inverse Problems*. ISTE, first edition, 2016.

Luo, G., Blumenthal, M., Heide, M., and Uecker, M. Bayesian MRI Reconstruction with Joint Uncertainty Estimation using Diffusion Models. *Magnetic Resonance in Medicine*, 90(1):295–311, 2023.

Martin, D., Fowlkes, C., Tal, D., and Malik, J. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, 2001.

Narnhofer, D., Habring, A., Holler, M., and Pock, T. Posterior-Variance-Based Error Quantification for Inverse Problems in Imaging. *ArXiv*, abs/2212.12499, 2022. URL https://arxiv.org/abs/2212.12499.

Peng, C., Guo, P., Zhou, S. K., Patel, V. M., and Chellappa, R. Towards Performant and Reliable Undersampled MR Reconstruction via Diffusion Model Sampling. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, 2022.

Ramzi, Z., Remy, B., Lanusse, F., Starck, J.-L., and Ciuciu, P. Denoising Score-Matching for Uncertainty Quantification in Inverse Problems. In *Proceedings of the International Conference on Neural Information Processing Systems*, 2020.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, 2015.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Proceedings of the International Conference on Neural Information Processing Systems*, 2019.

Song, Y. and Ermon, S. Improved Techniques for Training Score-Based Generative Models. In *Proceedings of the International Conference on Neural Information Processing Systems*, 2020.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In *Proceedings of the International Conference on Learning Representations*, 2021.

Szelsiki, R. *Computer Vision: Algorithms and Applications*. Springer, second edition, 2022.

Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., and Cattin, P. C. Diffusion Models for Implicit Image Segmentation ensembles. In *Proceedings of the International Conference on Medical Imaging with Deep Learning*, 2021.

Xie, Y. and Li, Q. Measurement-conditioned Denoising Diffusion Probabilistic Model for Under-sampled Medical Image Reconstruction. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, 2022.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B., and Yang, M.-H. Diffusion Models: A Comprehensive Survey of Methods and Applications. *ArXiv CoRR*, 2022. URL https://arxiv.org/abs/2209.00796.

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J., Defazio, A., Stern, R., Johnson, P., Bruno, M., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Yakubova, N., Pinkerton, J., Wang, D., Owens, E., Zitnick, C. L., Recht, M. P., Sodickson, D. K., and Lui, Y. W. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *ArXiv e-prints*, abs/1811.08839, 2018. URL https://arxiv.org/abs/1811.08839.

Acknowledgements

This project has received funding from the European Union's EIC Pathfinder Challenges 2022 programme under grant agreement No 101115317 (NEO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council. Neither the European Union nor the European Innovation Council can be held responsible for them.

Funded by
the European Union

A. Experiments

Experimental Settings. The BSDS consists of 400 training and test images and 100 validation images showing natural objects and humans. We use gray-scale images and augment the images by flipping them horizontally and vertically and cropping each of them five times into a patch size either of 64×64 or 96×96 pixels. The values of the variance schedule are based on those stated in Song & Ermon (2020) and Song et al. (2021) and are between $[0.1, 20]$. We use the U-Net (Ronneberger et al., 2015) with additional time step embeddings taken from Huang (2021) and modify it slightly for our purposes. All models are trained on a Nvidia Titan X GPU with 12 GB memory.

The fastMRI training set consists of 973 single-coil volumes containing 34,742 slices and the validation set consists of 199 volumes containing 7,135 slices (Zbontar et al., 2018). For the training process we use the single-coil knee coronal proton-density weighted images without fat suppression (PD images) and the images with fat suppression (PDFS images) and crop them randomly into a patch size of 64×64 or 96×96 pixels. As the k-space is complex also the image data is complex and is handled by using two channels, one for the real and one for the imaginary part. Instead of scaling the pixel value range of the images, we adjust the values for the variance schedule to $[1\text{e-}5, 1\text{e-}3]$. For the evaluation we only use a middle slice of the PD images as the PDFS images already contain a high noise level in the ground-truth data. Nevertheless, the approach also works with the PDFS images.

For the sampling process we either use the single or the multiple trajectories method. The multiple trajectories method takes the last sample of each reverse sampling process whereby the trajectories share the initial reverse steps. Contrary to that, the single trajectory method uses the samples of the last thousand steps of the reverse process whereby the variance of the added noise is constant for these steps. For the error quantification we compute the expected value of the posterior and the variance as well as the error to the ground-truth for the samples in the training set to obtain upper error bounds. For the validation set we calculate the variance and with the upper error bounds we are able to generate the predicted error quantiles.

Results. Table 1 and 2 summarize the results of the error quantification. The upper error bound guarantees for the 0.9 quantile are visualized in Figure 2 for the different scenarios. The original, corrupted and reconstructed images as well as the error and the estimated error quantile are visualized in Figure 3–6 for the different inverse problems. The results are represented for the models trained with the larger patch size and the multiple trajectories sampling procedure as these methods lead to improved results. The models trained with the smaller patch size tend to introduce artifacts which could be due to the lacking context in the tiny patches. Overall, we achieved superior outcomes with the models trained with larger image patches.

Table 1. The mutual information using the training set as well as the coverage of the 0.9 quantile and the PSNR values of the corrupted and reconstructed images using the validation set for the denoising (5% noise level) and inpainting (10% image information loss) task of the BSDS dataset.

Inverse Problem	Denoising		Inpainting
Method	Multiple Trajectories	Reference Narnhofer et al. (2022)	Multiple Trajectories
Mutual Information	0.1313	0.1231 - 0.1650	0.2494
Coverage 0.9 Quantile	0.8890	0.8951 - 0.9023	0.8839
PSNR Corrupted Image in dB	21.96	24.78	14.72
PSNR Reconstructed Image in dB	27.12	29.34 - 30.79	31.69

Table 2. The mutual information using the training set as well as the coverage of the 0.9 quantile and the PSNR values of the corrupted and reconstructed images using the validation set for the zero-filling tasks ($R = 4$ and $R = 8$) of the fastMRI dataset.

Acceleration Factor	$R = 4$		$R = 8$
Method	Multiple Trajectories	Reference Narnhofer et al. (2022)	Multiple Trajectories
Mutual Information	0.1233	0.1554	0.1203
Coverage 0.9 Quantile	0.8923	0.8966	0.8887
PSNR Corrupted Image in dB	20.69	10.66	19.92
PSNR Reconstructed Image in dB	29.75	34.79	28.04

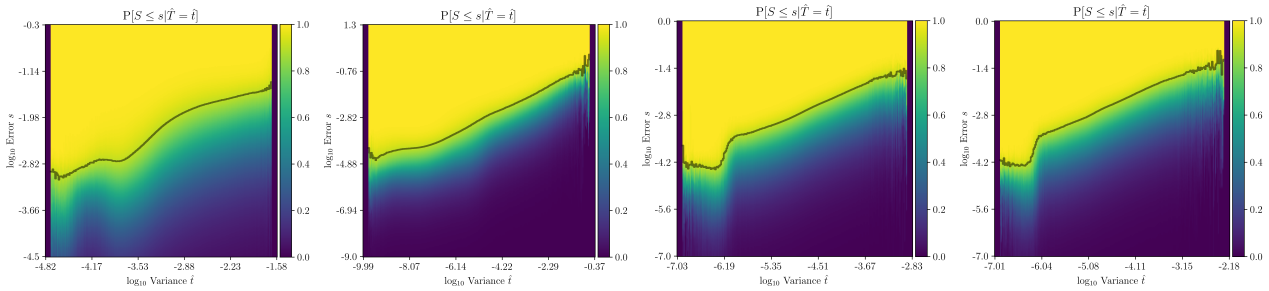

Figure 2. The cumulative conditional distribution $P[S \leq s|\hat{T} = \hat{t}]$ of the error s and the variance $\hat{t}$ as well as the 0.9 quantile of the denoising (5% noise), inpainting (10% image information loss) and the zero-filling task for $R = 4$ and $R = 8$ (from left to right).

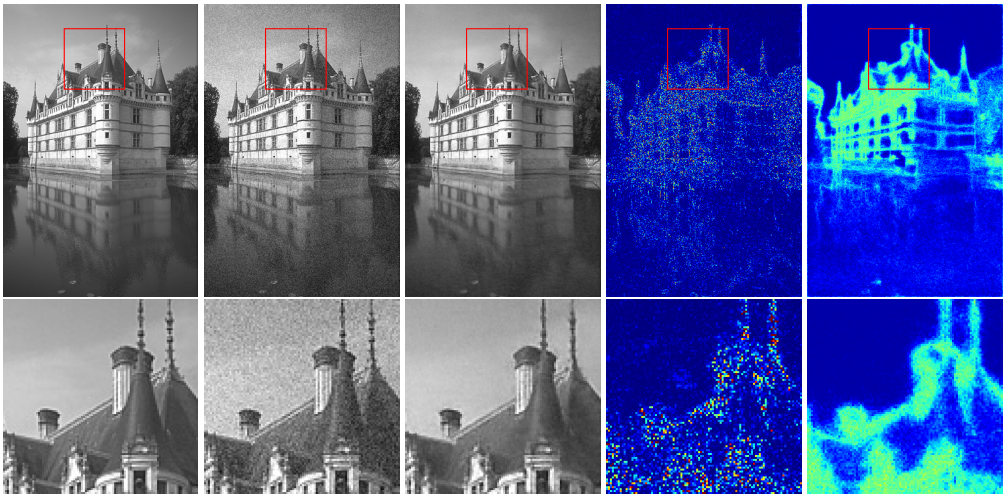

Figure 3. From left to right: The original, noisy and reconstructed images with the multiple trajectories method as well as the error and the 0.9 quantile images for the denoising task with 5% noise. The corresponding patches are in the second row. The range of the error and the quantile image is: 0 4e-2.

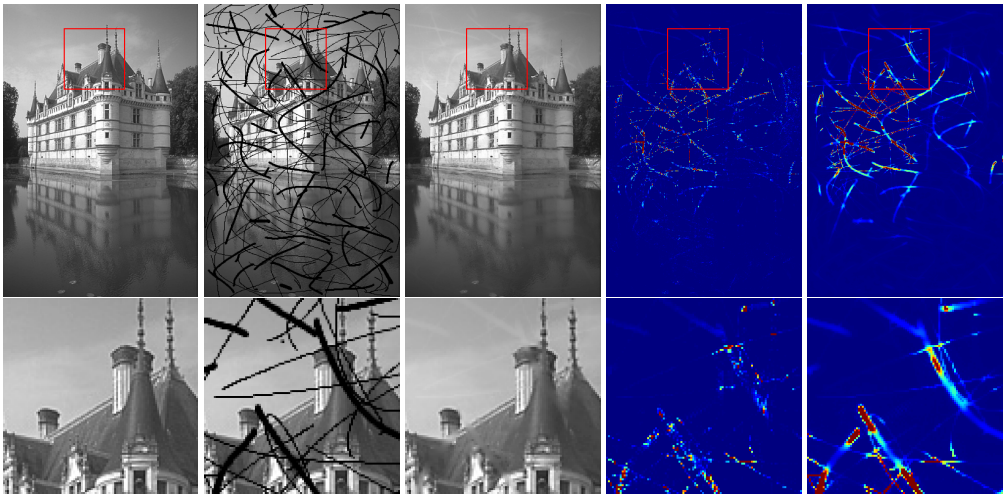

Figure 4. From left to right: The original, noisy and reconstructed images with the multiple trajectories method as well as the error and the 0.9 quantile images for the inpainting task with 10% image information loss. The corresponding patches are in the second row. The range of the error and the quantile image is: 0 4e-2.

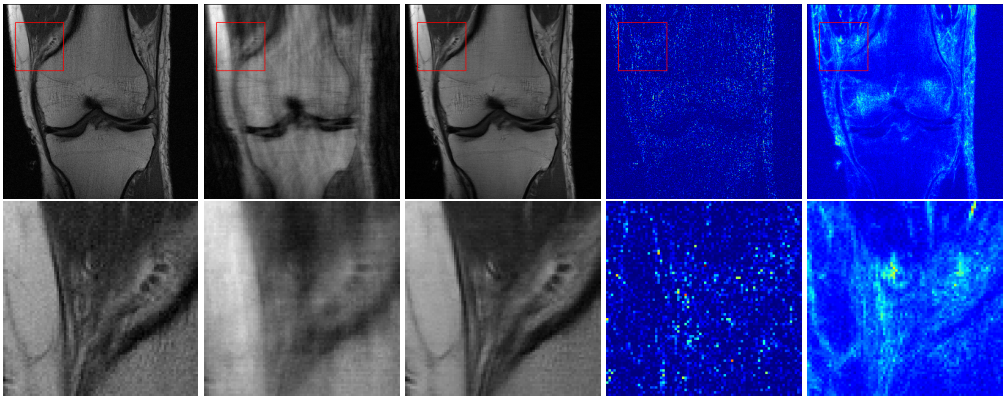

Figure 5. From left to right: The original, noisy and reconstructed images with the multiple trajectories method as well as the error and the 0.9 quantile images for the zero-filling task $R = 4$. The corresponding patches are in the second row. The range of the error and the quantile image is: 0 4e-2.

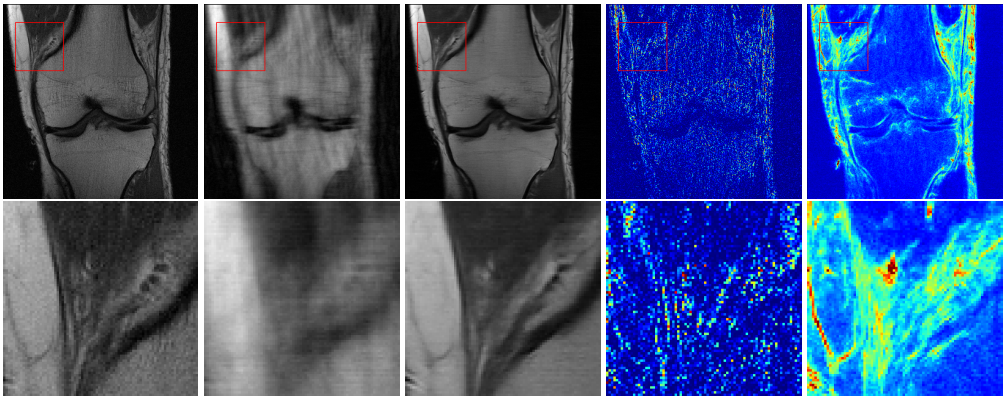

Figure 6. From left to right: The original, noisy and reconstructed images with the multiple trajectories method as well as the error and the 0.9 quantile images for the zero-filling task $R = 8$. The corresponding patches are in the second row. The range of the error and the quantile image is: 0 4e-2.