

# SCORE-BASED DIFFUSION MODELS FOR PHOTOACOUSTIC TOMOGRAPHY IMAGE RECONSTRUCTION

Sreemanti Dey, Snigdha Saha, Berthy T. Feng, Manxiu Cui, Laure Delisle,  
Oscar Leong, Lihong V. Wang, Katherine L. Bouman  
California Institute of Technology

## ABSTRACT

Photoacoustic tomography (PAT) is a rapidly-evolving medical imaging modality that combines optical absorption contrast with ultrasound imaging depth. One challenge in PAT is image reconstruction with inadequate acoustic signals due to limited sensor coverage or due to the density of the transducer array. Such cases call for solving an ill-posed inverse reconstruction problem. In this work, we use score-based diffusion models to solve the inverse problem of reconstructing an image from limited PAT measurements. The proposed approach allows us to incorporate an expressive prior learned by a diffusion model on simulated vessel structures while still being robust to varying transducer sparsity conditions.

Index Terms— Photoacoustic Tomography, Diffusion Models, Image Reconstruction, Generative Modeling

## 1. INTRODUCTION

Photoacoustic tomography (PAT) is a low-cost, ionizing-radiation-free technique for medical imaging. As such, it is growing in popularity and used in practical applications such as diagnosing breast cancer [1]. PAT measurements are sensor signals from a transducer array surrounding the object of interest, which then must be reconstructed into a human-interpretable image. However, physical and resource limitations may make it impossible to fully encompass the object with transducers (limited-view problem [2]) or build a dense-enough array to prevent aliasing (spatial-aliasing problem [3, 4]), limiting the reliability of a direct inversion.

With inadequate measurements, PAT image reconstruction can be formulated as an ill-posed inverse problem. Backprojection is a traditional solution but incorporates no priors and is prone to artifact-heavy reconstructions [5]. Model-based methods combine the measurement forward model with an image regularizer, but they do not capture complex image statistics, resulting in unrealistic reconstructions [6].

Deep learning poses an opportunity to incorporate more sophisticated image priors into the reconstruction. However, current deep-learning approaches are supervised with paired training data [7] and thus do not gen-

eralize to all measurement conditions. Practical applications call for a deep-learning approach that can be flexibly used in different settings.

Diffusion models are state-of-the-art generative models that have achieved success on various inverse imaging problems [8, 9, 10, 11, 12]. Song et al. [13] introduced a way to condition the generated images of a trained diffusion model on compressed-sensing measurements obtained for MRI or CT. However, PAT image reconstruction is not a compressed-sensing problem and instead involves dense, highly-correlated, time-varying measurements.

We introduce a method for PAT image reconstruction using a trained diffusion model. Our method is inspired by Song et al. [13] but generalizes to any type of linear inverse problem. We validate our approach on synthetic vascular structure images under different measurement conditions, including quantitative and qualitative comparisons to a supervised deep-learning method [7] and total-variation (TV) regularization [14]. Our work offers a technical contribution by proposing a new technique for solving general linear inverse problems with diffusion models, as well as a practical contribution by demonstrating the utility of diffusion models for PAT imaging.

## 2. BACKGROUND

### 2.1. Photoacoustic tomography (PAT)

PAT imaging maps optical absorption in scattering tissues with only surface-level measurements. In this work, we consider a ring-array-based PAT system for imaging a human breast. The breast of the patient is placed inside a ring of ultrasound sensors (Fig. 1). Short-pulsed laser light incident on the patient's skin diffuses deep into the breast tissue, and then the absorbed energy by blood vessels generates ultrasonic waves as a result of thermoelastic expansion. The ultrasonic waves are detected by the transducers. This spatio-temporal data is used to reconstruct the image of the object [15].

There are currently three major classes of image reconstruction methods for PAT: back-projection, time-reversal, and model-based. Back-projection methods do analytical inversion [16] and, while fast and tractable,

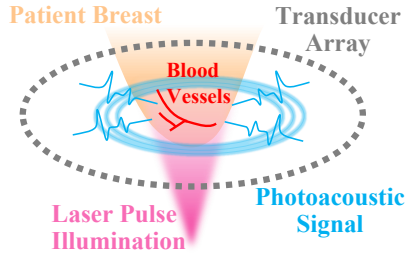


Fig. 1. PAT measurement acquisition. A ring of ultrasound sensors (transducer array) surrounds the object to be imaged. The transducer array receives photoacoustic signals emitted in response to a laser pulse.

often lead to images with artifacts [5]. Time-reversal methods, which use numerical simulations, give high-quality images but are computationally intensive [17]. Model-based methods minimize the difference between measured signals and predicted signals from an established forward model, often a linear operator [6]. Model-based methods are becoming more common due to their independence from measurement geometry and balance between computation and quality. Our work similarly uses a linear forward model based on curve-driven model matrix inversion (CDMMI) [6]. And while other deep-learning model-based methods exist (e.g., plug-and-play [18] and deep unrolling [19]), we leverage the strong prior of a diffusion model to achieve greater image quality. We note that concurrent work [20] applies diffusion models to PAT with a focus on the spatial aliasing problem.

## 2.2. Diffusion models for image reconstruction

Diffusion models are state-of-the-art generative models that learn to sample from an image distribution [21, 22, 23]. Recent methods have shown how to solve ill-posed inverse problems with a trained diffusion model as the prior [13, 8, 9, 11, 12], with most differing in the way measurements are incorporated into the sampling process of the diffusion model. Some of these methods have been applied to medical-imaging tasks like magnetic resonance imaging (MRI) [13, 24, 25], computed tomography (CT) [13], and ultrasound [26], but not to PAT. Our work builds upon an approach that has a simple projection step to incorporate measurements but was previously limited to compressed-sensing forward matrices [13]. By generalizing the projection step to any linear forward model, we are able to address PAT.

## 2.3. Score-based diffusion models

Diffusion models learn to sample from an image distribution through gradual denoising. Score-based diffusion models model the process of adding noise to an image as

a stochastic differential equation (SDE) [27]:

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t, \quad t \in [0, T] \quad (1)$$

where  $\mathbf{x}_t \in \mathbb{R}^d$  is the image;  $\mathbf{f}(\mathbf{x}_t, t)$  is the drift coefficient;  $g(t)$  is the diffusion coefficient; and  $d\mathbf{w}_t$  is infinitesimal white noise. This SDE gives rise to a time-dependent distribution  $p_t(\mathbf{x}_t)$ . Higher time  $t$  indicates more noise in  $\mathbf{x}_t$ . We specifically use the Variance-Preserving (VP) SDE [27] with  $T = 1$ , which ensures that  $p_T \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$ .

Sampling from the clean distribution  $p_0$  is based on the following reverse SDE:

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t, t) - g^2(t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)] dt + g(t)d\mathbf{w}_t. \quad (2)$$

Although the gradient  $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$  is unknown for an arbitrary image distribution  $p_0$ , it can be approximated with a convolutional neural network (CNN) called a score model  $\mathbf{s}_\theta$ :  $\mathbf{s}_\theta(\mathbf{x}, t) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ . The score model essentially learns to nudge images to higher probability.

Sampling starts with a noise image  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  that is gradually denoised by solving the reverse SDE (Eq. 2) with  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$  replaced by  $\mathbf{s}_\theta(\mathbf{x}, t)$ . Any numerical SDE solver can be used. We use a second-order solver via Predictor-Corrector sampling [27].

## 3. METHOD

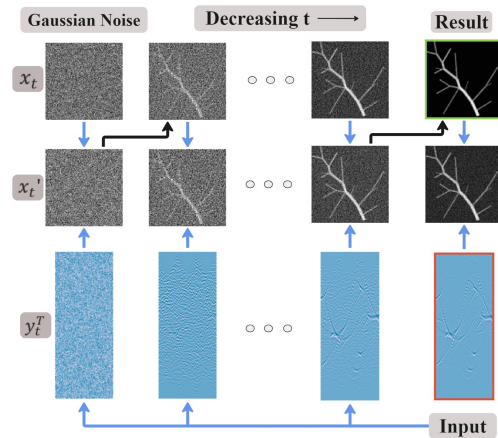


Fig. 2. Our conditional sampling process with a trained diffusion model. Given PAT measurements, sampling starts with an image of Gaussian noise, which is transformed over many steps into the reconstructed PAT image. Each step involves a measurement-conditioning update (blue arrow) followed by a denoising update (black arrow) that takes the image closer to the learned prior.

Our approach adapts the unconditional sampling procedure to be conditioned on PAT measurements  $\mathbf{y}$ . Following Song et al. [13], we model a diffusion process on  $\mathbf{y}$  and at each diffusion time  $t$ , modify the

image  $\mathbf{x}_t$  to be consistent with the perturbed measurements  $\mathbf{y}_t$ . Song et al. define the following measurement-conditioning step, which assumes an invertible matrix  $\mathbf{T}$  and a measurement-reduction operator  $\mathcal{P}(\Lambda)$ :

$$\mathbf{x}'_t \leftarrow \mathbf{T}^{-1}[\lambda\Lambda\mathcal{P}(\Lambda)\mathbf{y}_t + (1-\lambda)\Lambda\mathbf{T}\mathbf{x}_t + (\mathbf{I}-\Lambda)\mathbf{T}\mathbf{x}_t]. \quad (3)$$

Essentially,  $\mathbf{x}'_t$  balances the image  $\mathbf{x}_t$  produced by the unconditional diffusion model and the measurements  $\mathbf{y}_t$ , with  $\lambda \in [0, 1]$  tuning the weight of the measurements.

Our inverse problem, however, does not involve an invertible  $\mathbf{T}$  matrix or subsampling operator  $\mathcal{P}(\Lambda)$ . Instead, our CDMMI forward matrix [6] is an ill-conditioned tall matrix that produces highly-correlated measurements, making Eq. 3 unusable. We formulate a new measurement-conditioning step by solving a regularized maximum-likelihood objective:

$$\mathbf{x}'_t = \arg \min_{\mathbf{z} \in \mathbb{R}^d} [(1-\lambda)\|\mathbf{z} - \mathbf{x}_t\|_2^2 + \lambda\|\mathbf{y}_t - \mathbf{A}\mathbf{z}\|_2^2] \quad (4)$$

$$= (\lambda\mathbf{A}^\top\mathbf{A} + (1-\lambda)\mathbf{I})^{-1} ((1-\lambda)\mathbf{x}_t + \lambda\mathbf{A}^\top\mathbf{y}_t), \quad (5)$$

where  $\mathbf{A}$  is any forward matrix (in our case, the CDMMI matrix). Alg. 1 details our conditional sampling procedure, which is visualized in Fig. 2.

---

**Algorithm 1** Our conditional sampling process, where  $p_{0t}(\mathbf{y}_t | \mathbf{y})$  comes from the diffusion SDE (Eq. 1), and  $\Delta\mathbf{x}'_t$  is the SDE solver output at time  $t$  given  $\mathbf{x}'_t$ .

---

Require:  $N, T, \mathbf{y}, \mathbf{A}, \lambda$

$t \leftarrow T, \Delta t \leftarrow -\frac{T}{N}, \mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

while  $t > 0$  do

$\mathbf{y}_t \sim p_{0t}(\mathbf{y}_t | \mathbf{y})$

$\mathbf{x}'_t \leftarrow (\lambda\mathbf{A}^\top\mathbf{A} + (1-\lambda)\mathbf{I})^{-1} ((1-\lambda)\mathbf{x}_t + \lambda\mathbf{A}^\top\mathbf{y}_t)$

$\mathbf{x}_{t+\Delta t} \leftarrow \mathbf{x}'_t + \Delta\mathbf{x}'_t$

$t \leftarrow t + \Delta t$

end while

return  $\mathbf{x}_0$

---

## 4. EXPERIMENTS

We performed experiments on simulated measurements of both synthetic vascular images and a real breast-tissue image. We compared to two baseline methods: (1) maximum-likelihood with total-variation regularization (TV) and (2) a fully-supervised deep-learning approach called Sparse Artefact U-Net (SAU) [7]. Optimization for (1) was done with the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [28] [14]. (2) uses a U-Net CNN [29] to learn the error between naive reconstructions and ground-truth images—it is important to note that with this baseline, separate models must be trained for each transducer configuration. To create the paired training data, we simulated measurements of the ground-truth images and used Tikhonov-regularized MLE for naive reconstructions from these measurements.

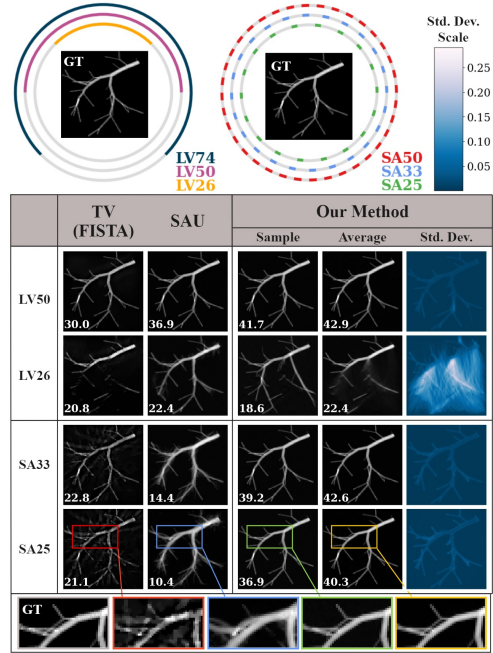


Fig. 3. Image reconstructions across meas. settings. Top two diagrams illustrate the limited-view and spatial-aliasing configurations, resp. (e.g., “LV74” refers to limited-view with 74% transducers). PSNR is on the bottom left of each image. Our method’s results include one sample and the avg. and std. dev. of 320 samples. The zoom-ins show high-fidelity details from our method that do not appear in baseline reconstructions. Overall, our method outperforms baselines in SA settings but may be prone to hallucination for LV (std. dev. maps show where hallucinations occur). Qualitatively, our samples tend to appear closer to the prior. The mean of our samples generally outperforms baselines.

We created a dataset of synthetic vascular structure images with Vascusynth [30], using the example parameters provided in the manual [31] but randomly setting the number of vascular nodes in each image. 9900 images were reserved for training our diffusion model and SAU, and 2000 images were reserved for validation of SAU.

We consider two types of limited sampling patterns of the transducer array: (1) limited view (LV) [2] and (2) spatial aliasing (SA) [3], illustrated in Fig. 3. LV is more challenging, as only a portion of the image circumference is observable. SA spaces the transducers equally around the circumference. We simulated measurements using a CDMMI forward matrix for each sampling configuration and added realistic Gaussian noise (30 dB SNR).

### 4.1. Image-reconstruction quality

In Fig. 4, we compare the average PSNR of reconstructions from our method versus those obtained with TV regularization and SAU. For SA, our method consistently

outperforms the baselines with a notable PSNR improvement (e.g., 2.4 to 23.3 dB improvement over SAU across all sparsity levels). For LV, SAU achieves slightly higher PSNRs, but our approach still averages within one SAU standard deviation. Fig. 3 shows reconstructions of an example test image.

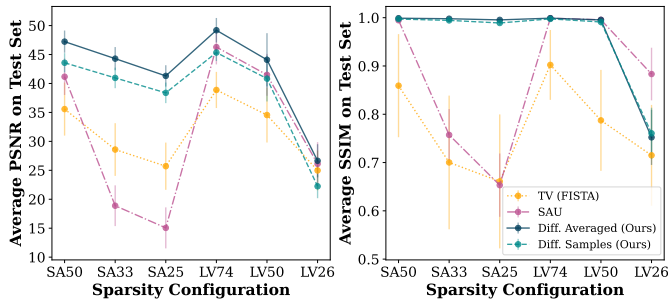


Fig. 4. Average PSNR and SSIM on the 10-image test set for TV, SAU, and our diffusion-model approach. “Diffusion Average” computes the avg. PSNR or SSIM based on the empirical mean of 320 samples; “Diffusion Sample” computes the avg. PSNR or SSIM based on all samples for each measurement. Our samples beat both baselines for SA configurations and perform on par for LV, while our averaged reconstructions outperform baselines on nearly every configuration.

The diffusion model excels at generating images true to its learned prior, but this also means that it may hallucinate structures when given very limited measurements. In particular, we find that our PSNR performance is worse than SAU in the LV setting due to hallucination. Although qualitatively our image samples are more visually-convincing, certain features in the image should be cautiously interpreted. One way to assess the reliability of reconstructed features is to compute the empirical standard deviation of many samples from the conditional sampling process (Alg. 1).

#### 4.2. Flexibility

We observe in Fig. 5 that transferring the same SAU to a different measurement setting results in significantly lower performance. In contrast, our method adapts to different settings without retraining.

Our approach’s flexibility also applies to out-of-distribution source images. Fig. 6 shows that our method can plausibly reconstruct a breast image [1] from simulated PAT measurements with good PSNR in all but the most extreme sparsity case, despite using a diffusion model trained only on synthetic vascular images.

### 5. DISCUSSION

We have presented a method for unsupervised PAT image reconstruction using a trained diffusion model. Our

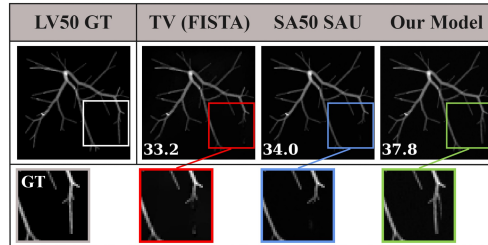


Fig. 5. Comparison of SAU performance on a new transducer pattern vs. our model and TV. Each model was tested on an LV50 image, but the SAU model used was trained on SA50 reconstructions. PSNRs show that SAU does not generalize across configurations. Zoom-in shows a GT feature that only our method was able to recover.

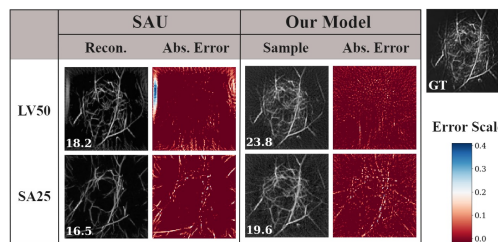


Fig. 6. Real breast image. Our diffusion model and SAU were each trained only on synthetic images. Ours produced higher-fidelity reconstructions for all configurations except LV26 (LV50 and SA25 shown here).

work builds upon a previous diffusion-model approach both by proposing a new measurement-conditioning formula suitable for any linear forward model and by tackling the problem of PAT imaging. In our experiments with simulated measurements, we find that our method performs substantially better than traditional TV regularization and competitively to a fully-supervised deep-learning approach (and even better when taking the sample mean), without requiring retraining for every transducer pattern. We also show better reconstruction of an out-of-distribution image of real breast tissue. Our work establishes a promising path to leveraging deep-learned priors for flexible photoacoustic tomographic imaging.

### 6. ACKNOWLEDGMENTS

This work was supported in part by a Heritage Medical Research Fellowship Award, National Institutes of Health grants U01 EB029823 (BRAIN Initiative), and R35 CA220436 (Outstanding Investigator Award). L.W. has a financial interest in Microphotoacoustics, Inc., CalPACT, LLC, and Union Photoacoustic Technologies, Ltd., which, however, did not support this work. B.T.F. is supported by the NSF GRFP. M.C. would like to thank Yousuf Aborahama for fruitful discussions.

## 7. REFERENCES

- [1] Li Lin, Peng Hu, Junhui Shi, Catherine M. Appleton, Konstantin Maslov, Lei Li, Ruiying Zhang, and Lihong V. Wang, “Single-breath-hold photoacoustic computed tomography of the breast,” *Nature Communications*, vol. 9, no. 1, pp. 2352, Jun 2018. [1](#), [4](#)
- [2] Yuan Xu, Lihong V Wang, Gaik Ambartsoumian, and Peter Kuchment, “Reconstructions in limited-view thermoacoustic tomography,” *Medical physics*, vol. 31, no. 4, pp. 724–733, 2004. [1](#), [3](#)
- [3] Peng Hu, Lei Li, Li Lin, and Lihong V Wang, “Spatiotemporal antialiasing in photoacoustic computed tomography,” *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3535–3547, 2020. [1](#), [3](#)
- [4] Peng Hu, Lei Li, and Lihong V Wang, “Location-dependent spatiotemporal antialiasing in photoacoustic computed tomography,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 4, pp. 1210–1224, 2022. [1](#)
- [5] Thomas Jetzfellner, Amir Rosenthal, K.-H. Englmeier, Alexander Dima, Miguel Ángel Caballero, Daniel Razansky, and Vasilis Ntziachristos, “Interpolated model-matrix optoacoustic tomography of the mouse brain,” *Applied Physics Letters*, vol. 98, no. 16, Apr 2011. [1](#), [2](#)
- [6] Hongbo Liu, Kun Wang, Dong Peng, Hui Li, Yukun Zhu, Shuang Zhang, Muhan Liu, and Jie Tian, “Curve-driven-based acoustic inversion for photoacoustic tomography,” *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2546–2557, 2016. [1](#), [2](#), [3](#)
- [7] Neda Davoudi, Xosé Luís Deán-Ben, and Daniel Razansky, “Deep learning optoacoustic tomography with sparse data,” *Nature Machine Intelligence*, vol. 1, no. 10, pp. 453–460, Oct 2019. [1](#), [3](#)
- [8] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song, “Denoising diffusion restoration models,” in *Advances in Neural Information Processing Systems*, 2022. [1](#), [2](#)
- [9] Shady Abu-Hussein, Tom Tirer, and Raja Giryes, “Adir: Adaptive diffusion for image reconstruction,” 2022. [1](#), [2](#)
- [10] Yinhuai Wang, Jiwen Yu, and Jian Zhang, “Zero-shot image restoration using denoising diffusion null-space model,” *arXiv preprint arXiv:2212.00490*, 2022. [1](#)
- [11] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye, “Diffusion posterior sampling for general noisy inverse problems,” 2023. [1](#), [2](#)
- [12] Berthy T Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T Freeman, “Score-based diffusion models as principled priors for inverse imaging,” *International Conference on Computer Vision (ICCV)*, 2023. [1](#), [2](#)
- [13] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon, “Solving inverse problems in medical imaging with score-based generative models,” 2022. [1](#), [2](#)
- [14] Amir Beck and Marc Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009. [1](#), [3](#)
- [15] Idan Steinberg, David M. Hulan, Ophir Vermesh, Hadas E. Frostig, Willemieke S. Tummers, and Sanjiv S. Gambhir, “Photoacoustic clinical imaging,” *Photoacoustics*, vol. 14, pp. 77–98, 2019. [1](#)
- [16] Minghua Xu and Lihong V Wang, “Universal back-projection algorithm for photoacoustic computed tomography,” *Physical Review E*, vol. 71, no. 1, pp. 016706, 2005. [1](#)
- [17] X. L. Dean-Ben, A. Buehler, V. Ntziachristos, and D. Razansky, “Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1922–1928, Oct 2012. [2](#)
- [18] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg, “Plug-and-play priors for model based reconstruction,” in *2013 IEEE global conference on signal and information processing*. IEEE, 2013, pp. 945–948. [2](#)
- [19] Vishal Monga, Yuelong Li, and Yonina C Eldar, “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing,” *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021. [2](#)
- [20] Xianlin Song, Guijun Wang, Wenhua Zhong, Kangjun Guo, Zilong Li, Xuan Liu, Jiaqing Dong, and Qiegen Liu, “Sparse-view reconstruction for photoacoustic tomography combining diffusion model with model-based iteration,” *Photoacoustics*, vol. 33, pp. 100558, 2023. [2](#)
- [21] Jonathan Ho, Ajay Jain, and Pieter Abbeel, “Denoising diffusion probabilistic models,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020. [2](#)
- [22] Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho, “Variational diffusion models,” *arXiv preprint arXiv:2107.00630*, 2021. [2](#)
- [23] Alexander Quinn Nichol and Prafulla Dhariwal, “Improved denoising diffusion probabilistic models,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 8162–8171. [2](#)
- [24] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jonathan I Tamir, “Robust compressed sensing mri with deep generative priors,” *NeurIPS*, 2021. [2](#)
- [25] Berthy T Feng and Katherine L Bouman, “Efficient bayesian computational imaging with a surrogate score-based prior,” *arXiv preprint arXiv:2309.01949*, 2023. [2](#)
- [26] Tristan SW Stevens, Faik C Meral, Jason Yu, Iason Z Apostolakis, Jean-Luc Robert, and Ruud JG van Sloun, “Dehazing ultrasound using diffusion models,” *arXiv preprint arXiv:2307.11204*, 2023. [2](#)
- [27] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, “Score-based generative modeling through stochastic differential equations,” 2021. [2](#)
- [28] Amir Beck and Marc Teboulle, “A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 693–696. [3](#)
- [29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer, 2015, pp. 234–241. [3](#)
- [30] Ghassan Hamarneh and Preet Jassi, “Vascusynth: Simulating vascular trees for generating volumetric image data with ground-truth segmentation and tree analysis,” *Computerized medical imaging and graphics*, vol. 34, no. 8, pp. 605–616, 2010. [3](#)
- [31] Preet Jassi and Ghassan Hamarneh, “Vascusynth: Vascular tree synthesis software,” *Insight Journal*, 2011. [3](#)