

NUCLEAR DIFFUSION MODELS FOR LOW-RANK BACKGROUND SUPPRESSION IN VIDEOS

Tristan S.W. Stevens*, Oisín Nolan*, Jean-Luc Robert[†] Ruud J.G. van Sloun*

*Dept. of Electrical Engineering, Eindhoven University of Technology, the Netherlands

[†]Philips Research North America, Cambridge MA, USA

ABSTRACT

Video sequences often contain structured noise and background artifacts that obscure dynamic content, posing challenges for accurate analysis and restoration. Robust principal component methods address this by decomposing data into low-rank and sparse components. Still, the sparsity assumption often fails to capture the rich variability present in real video data. To overcome this limitation, a hybrid framework that integrates low-rank temporal modeling with diffusion posterior sampling is proposed. The proposed method, *Nuclear Diffusion*, is evaluated on a real-world medical imaging problem, namely cardiac ultrasound dehazing, and demonstrates improved dehazing performance compared to traditional RPCA concerning contrast enhancement (gCNR) and signal preservation (KS statistic). These results highlight the potential of combining model-based temporal models with deep generative priors for high-fidelity video restoration.

Index Terms— RPCA, diffusion models, denoising

1. INTRODUCTION

Denoising, the recovery of a clean signal from a corrupted observation, is a foundational problem in signal processing [1], encompassing a diverse range of applications from natural image and video enhancement to sensory applications such as medical imaging and radar [2]. Typically, the objective is to disentangle informative structure from nuisance variability, thereby improving interpretability and downstream analysis of observed data. In sequential data, such as videos, the corruption is often structured rather than random, with background artifacts and stationary patterns that can obscure the dynamic content of interest.

Robust PCA (RPCA) provides a principled approach for separating foreground dynamics from structured background by decomposing sequential data into a *low-rank* background and a *sparse* component, representing the signal of interest. This idea has been widely applied in various domains, including speech [3], saliency detection, face recognition, and medical imaging [4]. While sparsity can be imposed in alternative transform domains (e.g., wavelet or frequency bases), such choices remain handcrafted and often insufficient to capture the rich variability of real signals.

Meanwhile, deep generative models (DGMs) are well-suited to capturing complex statistical dependencies, such as those present in videos. Diffusion models (DMs) [5, 6] in particular, have opened new directions in image restoration. DMs learn to sample from complex data distributions through iterative denoising, and have achieved state-of-the-art performance across inverse problems such as denoising, deblurring, and dehazing [7]. Acting as expressive learned priors, DMs can capture rich statistical structure beyond simple pixel-wise sparsity priors.

Building on this idea, we propose a hybrid denoising framework that combines the temporal modeling of RPCA with the expressive modeling of diffusion priors. In the proposed method, the conventional sparsity assumption on the foreground component is replaced with a learned diffusion prior, while maintaining a nuclear norm penalty to encourage low-rank temporal structure of the background. This is realized through diffusion posterior sampling (DPS) [8], which alternates between reverse diffusion and measurement-guided updates, ultimately allowing more accurate recovery of the dynamic foreground.

While our formulation is general-purpose, we evaluate our method on the real-world problem of cardiac ultrasound video dehazing, where structured noise (haze) degrades image quality and hampers diagnostic clarity [9]. RPCA has been a popular tool in ultrasound imaging, for example in clutter suppression [10] and elastography denoising [11], and has been extended through deep unfolding for high-dimensional settings [12]. Despite these advances, the sparsity assumption is often too restrictive in practice.

In this paper, we make the following contributions:

- A novel framework integrating DPS with a low-rank temporal model, generalizing RPCA via data-driven deep generative priors.
- An evaluation of the method on cardiac ultrasound dehazing, achieving enhanced image contrast while better preserving anatomical structures compared to standard RPCA.

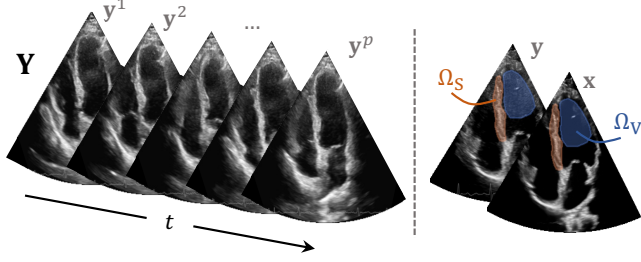


Fig. 1. Sequence of hazy cardiac ultrasound images $\mathbf{Y}$, with annotated regions of interest Ω used for evaluation.

2. BACKGROUND

2.1. Robust PCA for background suppression

Robust PCA (RPCA) decomposes observations $\mathbf{Y} \in \mathbb{R}^{n \times p}$ (e.g., pixel intensities of p frames, each of size n) into:

$$\mathbf{Y} = \mathbf{L} + \mathbf{X}, \quad (1)$$

where $\mathbf{L}$ is low-rank (coherent background, e.g., static haze) and $\mathbf{X}$ is sparse (foreground dynamics, e.g. tissue signal). Exact rank minimization is intractable, so RPCA solves the convex surrogate known as principal component pursuit (PCP). A common approach is to relax the hard equality constraint and form the Lagrangian:

$$\min_{\mathbf{L}, \mathbf{X}} \|\mathbf{L}\|_* + \lambda \|\mathbf{X}\|_1 + \frac{\mu}{2} \|\mathbf{Y} - \mathbf{L} - \mathbf{X}\|_F^2, \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and $\|\cdot\|_*$ denotes the nuclear norm, which serves as a convex surrogate for the rank of $\mathbf{L}$. In many imaging scenarios, the assumed signal model of (1) is overly simplistic. The signal component $\mathbf{X}$ often exhibits complex, structured patterns that are not truly sparse. As a result, standard RPCA tends to over-penalize these patterns, effectively attenuating or removing portions of the true signal. This limitation motivates replacing the generic ℓ_1 sparsity prior with a learned diffusion prior that better models the complex distribution of images, while retaining low-rank temporal modeling through a nuclear norm constraint.

2.2. Diffusion Posterior Sampling

Diffusion models learn the distribution $p(\mathbf{x})$ of a random variable $\mathbf{x}$ through a forward corruption process that gradually adds Gaussian noise:

$$\mathbf{x}_\tau = \alpha_\tau \mathbf{x}_0 + \sigma_\tau \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (3)$$

where $\mathbf{x}_0 \equiv \mathbf{x} \sim p(\mathbf{x})$, $\tau \in [0, \mathcal{T}]$, and α_τ, σ_τ are pre-defined noise schedules. The generative process is then defined as reversing this corruption, which is equivalent to iterative denoising of $\mathbf{x}_\tau$, starting from $\mathbf{x}_\mathcal{T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Tweedie's formula relates the minimum mean-square error estimate to the

Table 1. Comparison of RPCA and Nuclear Diffusion.

	RPCA	Nuclear Diffusion
$p(\mathbf{Y} \mathbf{L}, \mathbf{X})$	$\mathcal{N}(\mathbf{Y}; \mathbf{L} + \mathbf{X}, \sigma^2 \mathbf{I})$	$\mathcal{N}(\mathbf{Y}; \mathbf{L} + \mathbf{X}, \sigma^2 \mathbf{I})$
$p(\mathbf{L})$	$\propto \exp(-\ \mathbf{L}\ _*)$	$\propto \exp(-\ \mathbf{L}\ _*)$
$p(\mathbf{X})$	$\propto \exp(-\lambda \ \mathbf{X}\ _1)$	$p_\theta(\mathbf{X})$
Inference	$\arg \max_{\mathbf{L}, \mathbf{X}} p(\mathbf{L}, \mathbf{X} \mathbf{Y})$	$\mathbf{X}, \mathbf{L} \sim p_\theta(\mathbf{X}, \mathbf{L} \mathbf{Y})$

score of the distribution:

$$\mathbf{x}_{0|\tau} := \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_\tau] = \frac{1}{\alpha_\tau} (\mathbf{x}_\tau + \sigma_\tau^2 \nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)), \quad (4)$$

where $\mathbf{x}_{0|\tau}$ is a one-step denoised estimate, and the score $\nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)$ is parameterized by a neural network ϵ_θ , often predicting the noise, which relates to the score via $\epsilon_\theta(\mathbf{x}_\tau, \tau) \approx -\sigma_\tau \nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)$. To sample from the prior distribution $p(\mathbf{x})$, at each step τ , $\mathbf{x}_0$ is estimated using (4) and mapped back to $\mathbf{x}_{\tau-1}$ via forward diffusion in (3), ensuring a smooth sampling trajectory. The score network is trained using the denoising score matching objective [13]. *Unconditional* score-based diffusion models can be adapted for *conditional* sampling given noisy measurements, i.e., generating $\mathbf{x} \sim p(\mathbf{x} | \mathbf{y})$, by computing the score of the Bayesian posterior. Exact posterior sampling with diffusion models is generally intractable, but several approximate methods exist [7]. Here, we adopt diffusion posterior sampling (DPS) [8], which interleaves prior updates (denoising) with guidance steps (gradient steps toward the measurements). The prior update follows the same procedure as in unconditional sampling, while the guidance step, with forward model $\mathbf{y} = f(\mathbf{x}) + \mathbf{n}$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ is given by:

$$\nabla_{\mathbf{x}_\tau} \log p(\mathbf{y} | \mathbf{x}_\tau) \approx \nabla_{\mathbf{x}_\tau} \log p(\mathbf{y} | \mathbf{x}_{0|\tau}) \quad (5)$$

$$= -\frac{1}{2\sigma_n^2} \nabla_{\mathbf{x}_\tau} \|\mathbf{y} - f(\mathbf{x}_{0|\tau})\|_2^2. \quad (6)$$

We note that while $p(\mathbf{y} | \mathbf{x}_0)$ is often known exactly, the noise-perturbed likelihood $p(\mathbf{y} | \mathbf{x}_\tau)$ generally does not admit a closed form, motivating the approximation in (5).

3. METHODS

We adopt a Bayesian perspective to generalize the RPCA framework and extend it with a learned diffusion prior. Given observations $\mathbf{Y} \in \mathbb{R}^{n \times p}$ and independent latent variables $\mathbf{L}$ and $\mathbf{X}$ we construct the following joint distribution:

$$p(\mathbf{Y}, \mathbf{L}, \mathbf{X}) = p(\mathbf{Y} | \mathbf{L}, \mathbf{X}) p(\mathbf{L}) p(\mathbf{X}). \quad (7)$$

To arrive at the RPCA objective in (2), one can use a Gaussian forward model for the likelihood term:

$$p(\mathbf{Y} | \mathbf{L}, \mathbf{X}) = \mathcal{N}(\mathbf{Y}; \mathbf{L} + \mathbf{X}, \mu^{-1} \mathbf{I}). \quad (8)$$

Similarly, the low-rank component $\mathbf{L}$ follows a nuclear norm prior:

$$p(\mathbf{L}) \propto \exp(-\gamma \|\mathbf{L}\|_*), \quad (9)$$

which can be interpreted as a low-rank inducing prior, while the signal component $\mathbf{X}$ is modeled with a Laplace prior to enforce sparsity:

$$p(\mathbf{X}) \propto \exp(-\lambda \|\mathbf{X}\|_1). \quad (10)$$

Taking the negative logarithm of (7) and selecting the point estimate that maximizes the posterior (i.e., the MAP solution) recovers the classical RPCA objective in Lagrangian form, as shown in (2).

Algorithm 1 Nuclear Diffusion Posterior Sampling

Require: observations $\mathbf{Y}$, low-rank weighting γ , guidance weighting μ , diffusion model ϵ_θ , diffusion steps $\mathcal{T}$, noise schedule α_τ, σ_τ

- 1: Initialize $\mathbf{X}_\mathcal{T} \sim \mathcal{N}(\mathbf{0}, \sigma_\mathcal{T}^2 \mathbf{I})$, $\mathbf{L} \leftarrow \mathbf{0}$
 - 2: **for** $\tau = \mathcal{T}$ to 0 **do**
 - 3: $\epsilon^t \leftarrow \epsilon_\theta(\mathbf{x}_\tau^t, \tau)$, $\forall t = 1, \dots, p$ ▷ Predict noise
 - 4: $\mathbf{E}_\tau \leftarrow [\epsilon^1, \dots, \epsilon^p]$ ▷ Stack frames
 - 5: $\mathbf{X}_{0|\tau} \leftarrow \frac{1}{\alpha_\tau}(\mathbf{X}_\tau - \sigma_\tau \mathbf{E}_\tau)$ ▷ Denoise (prior)
 - 6: $\mathcal{E}_\tau \leftarrow \frac{\mu}{2} \|\mathbf{Y} - \mathbf{L} - \mathbf{X}_{0|\tau}\|_F^2$ ▷ Measurement error
 - 7: $\mathbf{X}_{0|\tau} \leftarrow \mathbf{X}_{0|\tau} - \nabla_{\mathbf{X}} \mathcal{E}_\tau$ ▷ Likelihood guidance
 - 8: $\mathbf{X}_{\tau-1} \leftarrow \alpha_\tau \mathbf{X}_{0|\tau} + \sigma_\tau \epsilon$ ▷ Forward diffusion
 - 9: $\mathcal{R}_\tau \leftarrow \gamma \|\mathbf{L}\|_*$ ▷ Low-rank penalty
 - 10: $\mathbf{L} \leftarrow \mathbf{L} - \nabla_{\mathbf{L}}(\mathcal{E}_\tau + \mathcal{R}_\tau)$ ▷ Background update
 - 11: **return** $\mathbf{X}_0, \mathbf{L}$
-

3.1. Nuclear diffusion

Building on this probabilistic formulation, we propose a hybrid framework that replaces the ℓ_1 sparsity prior on $\mathbf{X}$ with a learned diffusion prior $p_\theta(\mathbf{X})$ and performs posterior sampling instead of a MAP estimate, i.e., $\mathbf{X}, \mathbf{L} \sim p_\theta(\mathbf{X}, \mathbf{L} | \mathbf{Y})$. This allows $\mathbf{X}$ to capture complex, structured patterns beyond simple sparsity, while retaining the low-rank nuclear norm prior on $\mathbf{L}$ for temporal coherence. In practice, we implement this by interleaving a reverse diffusion process, as described in Section 2.2, with gradient-based guidance from the likelihood in (8) and the low-rank prior.

The diffusion prior $p_\theta(\mathbf{X})$ is applied in the spatial domain to individual frames, while temporal dependencies are enforced solely through the low-rank prior on $\mathbf{L}$. This separation enables the use of pretrained 2D diffusion models, simplifying implementation and avoiding the need for specialized video diffusion networks. Concretely, the signal component $\mathbf{X}$ can be written as:

$$\mathbf{X} = [\mathbf{x}^1 \quad \mathbf{x}^2 \quad \dots \quad \mathbf{x}^p] \in \mathbb{R}^{n \times p}, \quad (11)$$

where each column $\mathbf{x}^t \in \mathbb{R}^n$ represents the vectorized image at time t . The learned prior is parameterized by a denoising

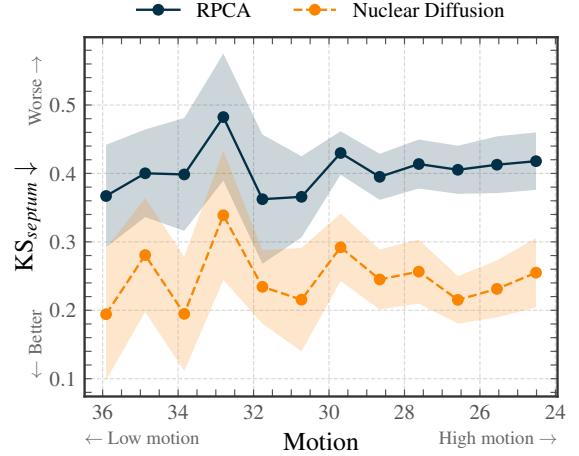


Fig. 2. KS statistic for various amount of motion levels measured via PSNR($\mathbf{y}^t, \mathbf{y}^{t-1}$), with Nuclear Diffusion outperforming RPCA across the entire range.

diffusion model $\epsilon_\theta(\mathbf{x}_\tau^t, \tau)$, which independently operates on a single noisy frame $\mathbf{x}_\tau^t$ at diffusion step τ :

$$\mathbf{x}_\tau^t \mapsto \epsilon_\theta(\mathbf{x}_\tau^t, \tau), \quad \forall t \in \{1, \dots, p\}. \quad (12)$$

Applying the per-frame denoiser independently across all frames induces a joint prior over the entire sequence, formalized in terms of scores for diffusion sampling [6]:

$$\nabla_{\mathbf{X}} \log p_\theta(\mathbf{X}) = -\frac{1}{\sigma_\tau} [\epsilon_\theta(\mathbf{x}_\tau^1, \tau), \dots, \epsilon_\theta(\mathbf{x}_\tau^p, \tau)]. \quad (13)$$

The inference framework is detailed in Algorithm 1 and a comparison of the distributions is given in Table 1.

4. RESULTS

We evaluate the proposed method on the task of cardiac ultrasound dehazing, focusing on both haze removal and tissue structure preservation. Given that a ground truth is not available, performance is assessed using two unsupervised metrics: generalized contrast-to-noise ratio (gCNR) [14], which we use to measure contrast between ventricle Ω_V and septum Ω_S regions. Additionally, we use the Kolmogorov–Smirnov (KS) statistic to quantify agreement between the original $\mathbf{Y}$ and denoised $\mathbf{X}$ tissue distributions in the septum region Ω_S .

$$\text{KS} = \sup_z \left| F_{\Omega_S(x)}(z) - F_{\Omega_S(y)}(z) \right|, \quad (14)$$

where $F(\cdot)$ is the empirical CDF of the respective ROIs. See Fig. 1 for an example sequence with annotated regions. The dataset contains videos of $60 \times 256 \times 256$, with in total 4,376 clean frames from 75 easy-to-image subjects and 2,324 noisy frames from 40 difficult-to-image subjects [15]. Fig. 3 presents a qualitative comparison. Both Nuclear Diffusion

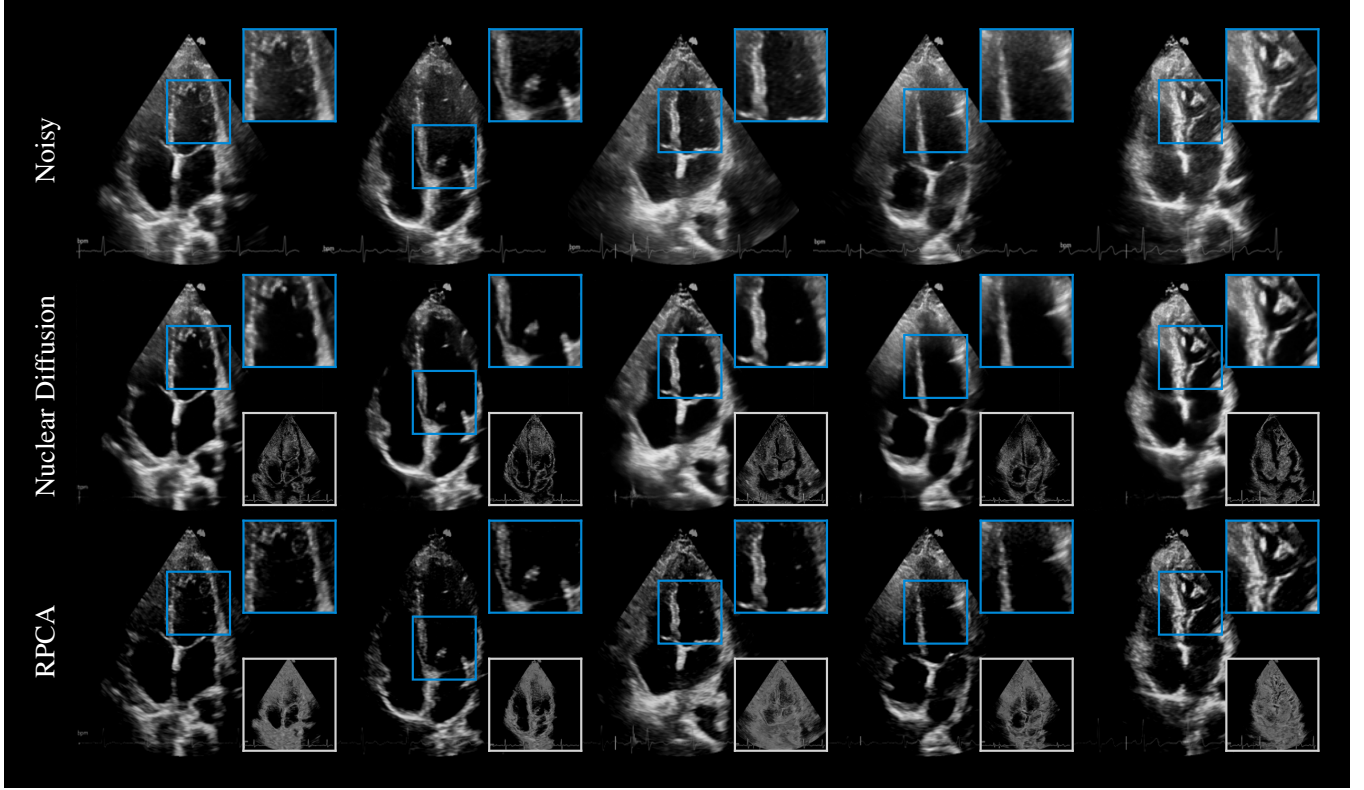


Fig. 3. Comparison on the task of cardiac ultrasound dehazing. While both methods suppress haze (shown in bottom insets), RPCA tends to excessively attenuate tissue, resulting in sparse structures, whereas Nuclear Diffusion better preserves details.

and RPCA reduce haze, but RPCA over-attenuates tissue, resulting in sparse or discontinuous structures, while Nuclear Diffusion adheres better to the prior trained on the clean dataset. Quantitative results support these observations, as shown in Fig. 4 and Fig. 2. Results are generated with 500 diffusion steps, accelerated using SeqDiff [16] ($T = 5000$), with Y^t as initialization. Furthermore, we use $\gamma = 1$, $\mu = 2$, and $p = 7$. The method is implemented using the *zea* [17] library with JAX backend.

5. CONCLUSIONS

In this paper, we introduced a hybrid framework that generalizes RPCA by integrating low-rank temporal modeling with learned generative diffusion priors. By replacing the standard ℓ_1 sparsity prior with a score-based generative model and performing diffusion posterior sampling with a nuclear norm penalty, our approach captures complex signal components while explicitly separating dynamic foreground from low-rank background. We demonstrated the effectiveness of this method on cardiac ultrasound video dehazing, showing that it can suppress haze artifacts and improve image contrast while preserving delicate tissue. These results highlight the potential of combining classical low-rank priors with modern generative models for video restoration.

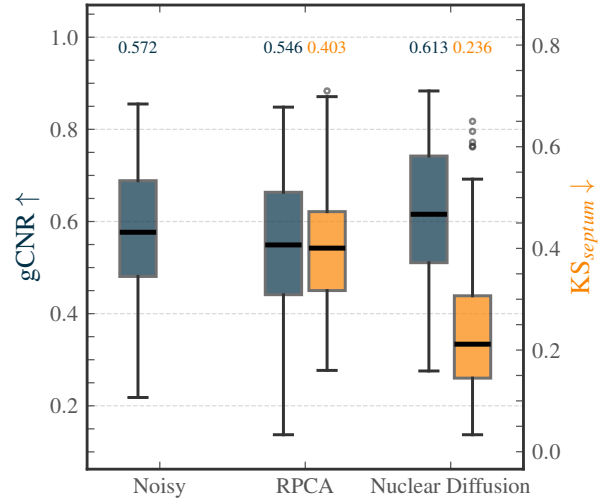


Fig. 4. Quantitative comparison of Nuclear Diffusion and RPCA using gCNR and KS metrics. Nuclear Diffusion achieves higher contrast between Ω_S and Ω_V while preserving the tissue intensity distribution in Ω_S , whereas RPCA tends to attenuate tissue and distort signal statistics.

6. REFERENCES

- [1] Peyman Milanfar and Mauricio Delbracio, “Denoising: A Powerful Building-block for Imaging, Inverse Problems, and Machine Learning,” *Philosophical Transactions A*, vol. 383, no. 2299, pp. 20240326, 2025.
- [2] Tristan S W Stevens, Jeroen Overdevest, Oisín Nolan, Wessel L van Nierop, Ruud J G van Sloun, and Yonina C Eldar, “Deep generative models for bayesian inference on high-rate sensor data: applications in automotive radar and medical imaging,” *Philos. Trans. A Math. Phys. Eng. Sci.*, vol. 383, no. 2299, pp. 20240327, 2025.
- [3] Mohaddeseh Mirbeygi, Aminollah Mahabadi, and Akbar Ranjbar, “Rpga-based real-time speech and music separation method,” *Speech Communication*, vol. 126, pp. 22–34, 2021.
- [4] Thierry Bouwmans, Sajid Javed, Hongyang Zhang, Zhouchen Lin, and Ricardo Otazo, “On the applications of robust pca in image and video processing,” *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1427–1457, 2018.
- [5] Jonathan Ho, Ajay Jain, and Pieter Abbeel, “Denoising Diffusion Probabilistic Models,” in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, Eds., 2020.
- [6] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, “Score-based Generative Modeling through Stochastic Differential Equations,” in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. 2021, OpenReview.net.
- [7] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G. Dimakis, and Mauricio Delbracio, “A Survey on Diffusion Models for Inverse Problems,” *CoRR*, vol. abs/2410.00083, 2024.
- [8] Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye, “Diffusion Posterior Sampling for General Noisy Inverse Problems,” in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. 2023, OpenReview.net.
- [9] Tristan S. W. Stevens, Faik C. Meral, Jason Yu, Iason Zacharias Apostolakis, Jean-Luc Robert, and Ruud J. G. van Sloun, “Dehazing Ultrasound Using Diffusion Models,” *IEEE Trans. Medical Imaging*, vol. 43, no. 10, pp. 3546–3558, 2024.
- [10] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud JG van Sloun, and Yonina C Eldar, “Deep unfolded robust pca with application to clutter suppression in ultrasound,” *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1051–1063, 2019.
- [11] Md Ashikuzzaman and Hassan Rivaz, “Denoising rf data via robust principal component analysis: Results in ultrasound elastography,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2067–2070.
- [12] HanQin Cai, Jialin Liu, and Wotao Yin, “Learned robust pca: A scalable deep unfolding approach for high-dimensional outlier detection,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 16977–16989, 2021.
- [13] Pascal Vincent, “A Connection Between Score Matching and Denoising Autoencoders,” *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, July 2011, Conference Name: Neural Computation.
- [14] Alfonso Rodriguez-Molares, Ole Marius Hoel Rindal, Jan D’hooge, Svein-Erik Måsøy, Andreas Austeng, Muyinatu A Lediju Bell, and Hans Torp, “The generalized contrast-to-noise ratio: A formal definition for lesion detectability,” *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 4, pp. 745–759, 2019.
- [15] Yi Guo, Yuanyuan Wang, Zeju Li, Jing Jiao, Xue Gao, Yunshu Li, Wei Guo, He Li, and Xiaozhou Zhou, “Dehazing echocardiography challenge 2025,” <https://dehazingecho2025.grand-challenge.org/>, 2025, Grand Challenge, MICCAI 2025.
- [16] Tristan S. W. Stevens, Oisín Nolan, Jean-Luc Robert, and Ruud J. G. van Sloun, “Sequential Posterior Sampling with Diffusion Models,” in *2025 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2025, Hyderabad, India, April 6-11, 2025*. 2025, pp. 1–5, IEEE.
- [17] Tristan S.W. Stevens, Wessel L. van Nierop, Ben Luijten, Vincent van de Schaft, Oisín I. Nolan, Beatrice Federici, Louis D. van Harten, Simon W. Penninga, Noortje I.P. Schueler, and Ruud J.G. van Sloun, “zea: A Toolbox for Cognitive Ultrasound Imaging,” July 2025.