# Deformed2Self: Self-supervised Denoising for Dynamic Medical Imaging

Junshen Xu[1](✉) and Elfar Adalsteinsson[1,2]

[1] Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA
junshen@mit.edu

[2] Institute for Medical Engineering and Science, MIT, Cambridge, MA, USA

**Abstract.** Image denoising is of great importance for medical imaging system, since it can improve image quality for disease diagnosis and downstream image analyses. In a variety of applications, dynamic imaging techniques are utilized to capture the time-varying features of the subject, where multiple images are acquired for the same subject at different time points. Although signal-to-noise ratio of each time frame is usually limited by the short acquisition time, the correlation among different time frames can be exploited to improve denoising results with shared information across time frames. With the success of neural networks in computer vision, supervised deep learning methods show prominent performance in single-image denoising, which rely on large datasets with clean-vs-noisy image pairs. Recently, several self-supervised deep denoising models have been proposed, achieving promising results without needing the pairwise ground truth of clean images. In the field of multi-image denoising, however, very few works have been done on extracting correlated information from multiple slices for denoising using self-supervised deep learning methods. In this work, we propose Deformed2Self, an end-to-end self-supervised deep learning framework for dynamic imaging denoising. It combines single-image and multi-image denoising to improve image quality and use a spatial transformer network to model motion between different slices. Further, it only requires a single noisy image with a few auxiliary observations at different time frames for training and inference. Evaluations on phantom and *in vivo* data with different noise statistics show that our method has comparable performance to other state-of-the-art unsupervised or self-supervised denoising methods and outperforms under high noise levels.

**Keywords:** Image denoising · Dynamic imaging · Deep learning · Self-supervised learning

**Electronic supplementary material** The online version of this chapter (https://doi.org/10.1007/978-3-030-87196-3_3) contains supplementary material, which is available to authorized users.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12902, pp. 25–35, 2021.
https://doi.org/10.1007/978-3-030-87196-3_3

## 1 Introduction

Noise is inevitable in medical images. A variety of sources lead to noisy images, such as acquisition with better spatial resolution in MRI [12] and reduction of radiation dose PET [30] and CT [27]. Further complicating the task is the complex noise statistics in different medical imaging modalities, which is not limited to additive white Gaussian noise. For example, the noise can be related to pixel intensity like Rician noise in magnitude images of MRI [31] or it can be affected by geometry parameters of the scanner like in CT [23]. Thus, a robust method for noise reduction plays an important role in medical image processing and also serves as a core module in many downstream analyses.

In many applications, more than one image is acquired during the scan to capture the dynamic of the subject, e.g., cine images for cardiac MRI [20], abdominal dynamic contrast-enhanced (DCE) MRI [11], and treatment planning 4D thoracic CT [4]. Image denoising is more necessary for dynamic imaging methods, since they often adopt fast imaging techniques to improve temporal resolution, which may reduce signal-to-noise ratio of each time frame. Dynamic imaging also provides more information for denoising as the images at different time frames have similar content and often follow the same noise model.

In recent years, a number of deep learning based methods have been proposed for image denoising [6,10,15,32]. Similar methods are also applied to medical image denoising, e.g., low-dose PET [29], CT [5] and MRI [28] denoising. These methods train a convolution neural network (CNN) that maps the noisy image to its clean counterpart. However the supervised training process requires a large-scale dataset with clean and noisy image pairs which can be difficult to acquire, especially in the field of medical imaging. Recently, there are several studies on learning a denoising network with only noisy images. The Noise2Noise [16] method trains the model on pairs of noisy images with the same content but different noises so that the network estimates the expectation of the underlaying clean image. However, in practice, it is difficult to collect a set of different noisy realizations of the same image. The Noise2Void [14] and Noise2Self [2] methods try to learn denoising network using a dataset of unpaired noisy images and use the blind-spot strategy to avoid learning a identity mapping. Yet, to achieve good performance, these two methods still require the test images to be similar to the training images in terms of image content and noise model. To address this problem, some denoising methods have been developed that only train networks with internal information from a single noisy image and do not rely on large training set. Ulyanov *et al.* introduced deep image prior (DIP) for single-image recovery [26]. Recently, Quan *et al.* proposed Self2Self [21] method, where it uses dropout [25] to implement the blind-spot strategy.

In terms of denosing methods for dynamic imaging, Benou *et al.* proposed a spatio-temporal denoising network for DCE MRI of brain, where the motion is negligible. Another category of methods first apply conventional registration method to register the images and then perform denoising on the registered images [18,23]. However, traditional optimization methods are time consuming [1] and registering noisy images directly may reduce the accuracy of registration.

In this work, we propose a deep learning framework for dynamic imaging denoising, named Deformed2Self, where we explore similarity of images in dynamic imaging by deforming images at different time frames to the target frame and utilize the fact that noises of different observations are independent and following similar noise model. Our method has the following features: 1) The whole pipeline can be trained end-to-end, which is efficient for optimization. 2) Our method is fully self-supervised, i.e., we only need noisy images without ground-truth clean images. 3) The model can be trained on a single image (with a few auxiliary observations) and has no prerequisite on large training dataset, making it suitable for applications with scarce data.

## 2 Methods

### 2.1 Problem Formulation

Let $y_0$ be the noisy image we want to denoise, which is generated with some noise models, e.g., for additive noise,

$$y_0 = x_0 + n_0, \tag{1}$$

where $x_0$ denotes the unknown clean image, and $n_0$ denotes the random measurement noise. The goal of single-image denoising is to find a mapping $f$ that can recover $x_0$ from $y_0$, i.e., $\hat{x}_0 = f(y_0) \approx x_0$.

In dynamic imaging, multiple images are acquired for the same subject at different time frames. Suppose we have another $N$ frames besides the target frame $y_0$. The noisy observations of these $N$ frames and their unknown clean counterparts are denoted as $\{y_1, ..., y_N\}$ and $\{x_1, ..., x_N\}$ respectively, where $y_k = x_k + n_k$, $k = 1, ..., N$. For dynamic imaging denoising, information from different time frames are aggregated to estimate the clean image at target frame (frame 0), $\hat{x}_0 = f(y_0, y_1, ..., y_N)$. In many cases, the motion of subject occur during the scan is not negligible. Let $\phi_k$ be the deformation field between the target frame and frame $k$, $x_0 = x_k \circ \phi_k$. The noisy observation $y_k$ can be rewritten as

$$y_k = x_k + n_k = x_0 \circ \phi_k^{-1} + n_k, \quad k = 0, 1, ..., N. \tag{2}$$

where $\phi_0 = \mathbf{I}$ is the identity mapping. Equation 2 indicates that the observed noisy images $\{y_k\}_{k=0}^{N}$ is generated from the target clean image $x_0$ following certain motion models and noise statistics. Therefore the auxiliary images from different time frames provide information for estimating the clean image at target frame.

### 2.2 The Deformed2Self Framework

Inspired by the data model above, we proposed a self-supervised denoising framework for dynamic imaging named Deformed2Self (Fig. 1), which consists of three

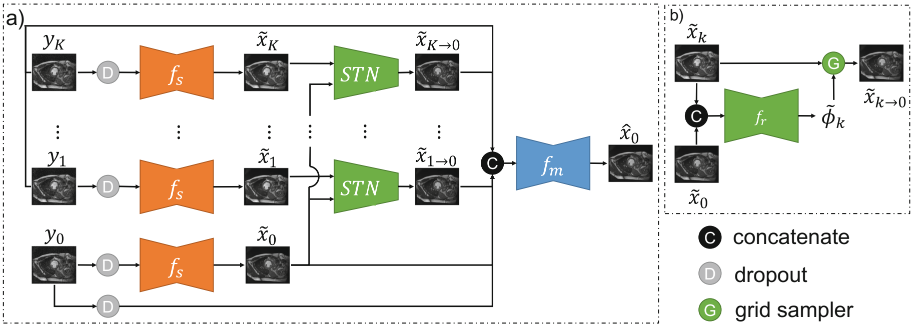

**Fig. 1.** Overview of the proposed self-supervised denoising framework. a) The architecture of Deformed2Self, where $f_s$ and $f_m$ are single- and multi-image denoising networks respectively. b) Details of the STN module, where $f_r$ is a registration network.

modules: 1) a single-imaging denoising network to provide coarse estimation of the clean image, 2) a spatial transformer network (STN) [9] for matching spatial content of images from other time frames ($k = 1, ..., N$) to the target frame, and 3) a multi-image denoising network to generate a refined estimation of the clean image at target frame.

**Single-Image Denoising:** The first step of the proposed framework is single-image denoising, where we use a UNet-based [22] network $f_s$ to denoise each frame separately. The benefit of this step is two-fold: 1) It provides a coarse estimation of the clean image utilizing only the information of each frame, which serves as a good initialization for the multi-image denoising to accelerate convergence. 2) It also improves the registration accuracy in the following STN module, since the prediction of deformation field suffers heavily from image noise.

To denoise images with internal information of each frame, we adopt the dropout-based blind-spot strategy [14,21]. Given a noisy observation $y_k$, $k \in \{0, ..., N\}$, we use a dropout layer to mask out some pixels in $y_k$ before feeding it to the network. The output can be written as $\tilde{x}_k = f_s(b_k \odot y_k)$, where $b_k$ is a random binary mask and $\odot$ is the Hadamard product. The network $f_s$ is trained to recover the missing pixels based on information from the remaining pixels. The network's best guess would be the expected pixel value [14], and therefore the output image $\tilde{x}_k$ can be considered as the denoised version of $y_k$.

The network $f_s$ is trained by minimizing the following loss function, which is the mean squared error (MSE) between the network output and the noisy image on the masked pixels.

$$\mathcal{L}_s = \frac{1}{N+1} \sum_{k=0}^{N} ||(1 - b_k) \odot (\tilde{x}_k - y_k)||_2^2 \tag{3}$$

**Spatial Transformer Network:** Let $\tilde{n}_k$ be the residual between the denoised image $\tilde{x}_k$ and the underlying clean image $x_k$, i.e., $\tilde{n}_k = \tilde{x}_k - x_k$, and thus

$\tilde{x}_k = x_0 \circ \phi_k^{-1} + \tilde{n}_k$. Suppose we can estimate $\phi_k$, the deformation field between frame $k$ and the target frame, and apply it to the denoised image $x_k$, then $\tilde{x}_k \circ \phi_k = x_0 + \tilde{n}_k \circ \phi_k = x_0 + \tilde{n}'_k$, i.e., $\tilde{x}_k \circ \phi_k$ is an image that is spatially matched with $x_0$ but is corrupted by noise $\tilde{n}'_k = \tilde{n}_k \circ \phi_k$. Note that $\{\tilde{x}_0, \tilde{x}_1 \circ \phi_1, ..., \tilde{x}_N \circ \phi_N\}$ can be considered as a set of images that share the same spatial content but have different noise and can be used for multi-image denoising later.

We estimate the motion between the target frame and other frames using a STN module (Fig. 1b). A network $f_r$ is used to predict the deformation field given pairs of moving image $\tilde{x}_k$ and target image $\tilde{x}_0$, $\tilde{\phi}_k = f_r(\tilde{x}_k, \tilde{x}_0)$. Then $\tilde{x}_k$ is deformed with a grid sampler, $\tilde{x}_{k\to 0} = \tilde{x}_k \circ \tilde{\phi}_k$. We adopt the architecture in [1] for $f_r$ and optimize it to minimize the following loss function,

$$\mathcal{L}_r = \frac{1}{N} \sum_{k=1}^{N} ||\tilde{x}_{k\to 0} - \tilde{x}_0||_2^2 + \lambda ||\nabla \tilde{\phi}_k||_2^2, \tag{4}$$

where $\lambda$ is a weighting coefficient. The first term is the MSE between the warped image $\tilde{x}_k \circ \tilde{\phi}_k$ and the target image $\tilde{x}_0$, which is the similarity metric. The second term is the L2 norm of spatial gradient of $\tilde{\phi}_k$, serving as a regularization for the deformation field [1].

**Multi-image Denoising:** We now have two sets of images, the denoised and deformed images $\{\tilde{x}_0, \tilde{x}_{1\to 0}, ..., \tilde{x}_{N\to 0}\}$ and the original noisy images $\{y_0, y_1, ..., y_N\}$. In the final stage, we aggregate all these images to generate a refined estimation of the target clean image. We adopt the blind-spot method similar to the single-image denoising stage but concatenate all images as input and produce an estimation for the target frame, $\hat{x}_0 = f_m(\tilde{x}_0, \tilde{x}_{1\to 0}, ..., \tilde{x}_{N\to 0}, b \odot y_0, y_1, ..., y_N)$, where $\hat{x}_0$ is the final estimation and $f_m$ is the multi-image denoising network. Again, we use dropout to remove some pixels in $y_0$ to avoid learning a identity mapping. $f_m$ shares the same architecture with $f_s$ except for the number of input channels. Similar to Eq. 3, a masked MSE loss is used to train network $f_m$,

$$\mathcal{L}_m = ||(1 - b) \odot (\hat{x}_0 - y_0)||_2^2. \tag{5}$$

### 2.3 Training and Inference

As mentioned above, all the three modules in the proposed method is trained in unsupervised or self-supervised manners. Besides, we can train this pipeline end-to-end with gradient descent based methods. Another advantage of our framework is that it can be trained on a single noisy image (with several auxiliary observations) without large-scale external dataset. In summary, the total loss function for our proposed model is $\mathcal{L} = \lambda_s \mathcal{L}_s + \lambda_r \mathcal{L}_r + \mathcal{L}_m$, where $\lambda_s$ and $\lambda_r$ are weighting coefficients. During training, the networks are updated for $N_{train}$ iterations to learning a specific denoising model for the input images. At each iteration, input images are randomly rotated for data augmentation and the dropout layers sample different realizations of binary masks $\{b, b_0, ..., b_N\}$ so that the denoising networks can learn from different parts of the image.

For inference, we run forward pass of the model for $N_{test}$ times with dropout layers enabled and average output $\hat{x}_0$'s to generate the final prediction.

## 3 Experiments and Results

### 3.1 Datasets and Experiments Setup

We use the following two datasets to evaluate our proposed denoising framework.

**PINCAT:** The PINCAT phantom [17,24] simulates both cardiac perfusion dynamics and respiration with variability in breathing motion. The spatial matrix size of the phantom is 128 × 128, which corresponds to a resolution of 1.5 mm × 1.5 mm. The intensity of the phantom is normalized to $[0, 1]$. The middle frame of the series is selected as the target frame.

**ACDC:** The ACDC dataset [3] consists of short-axis cardiac cine MRI of 100 subjects. The in-plane resolution and gap between slices range from 0.7 mm × 0.7 mm to 1.92 mm × 1.92 mm and 5 mm to 10 mm respectively. For pre-processing, we linearly normalize the image intensity to $[0, 1]$ and crop or pad them to a fixed size of 224 × 224. For each subject, we extract the sequence of the middle slice and use the end-systole (ES) and end-diastole (ED) phases as target frames.

In all the experiments, we use another four noisy images from the same sequences as auxiliary observations (two adjacent frames before and after the target frame), i.e., $N = 4$. We set $\lambda = 0.1$, and $\lambda_s = \lambda_r = 1$ in the loss function. For dropout layers, a dropout rate of 0.3 is used. We train the model using an Adam optimizer [13] with a learning rate of $1 \times 10^{-4}$. $N_{train}$ is set to 2000 for PINCAT dataset and 4000 for ACDC dataset. $N_{test}$ is set to 100 for both datasets. The neural networks are implemented with PyTorch 1.5, and trained and evaluated on a Titan V GPU. For PINCAT dataset, we simulate Gaussian noise and Poisson noise at different noise levels. The standard deviation $\sigma$ of Gaussian noise is set to 15%, 20% and 25%. The noisy observation $y$ under Poisson noise is generated by $y = z/P$, where $z \sim \text{Pois}(Px)$, $x$ is the truth intensity and $P$ is a parameter to control noise level. We set $P = 40, 20$, and 10 in the experiments. For ACDC dataset, we simulate Gaussian noise and Rician noise [8], with $\sigma = 5\%, 10\%$, and 15%. We compare our proposed method (D2S) with other state-of-the-art denoising methods, including deep image prior (DIP) [26], Self2Self (S2S) [21], BM3D [7] and VBM4D [19]. For DIP and S2S, we use the same learning rate mentioned above and tune the number of iterations on our datasets. We adopt the peak signal to noise ratio (PSNR) and the structural similarity index measure (SSIM) as evaluation metrics. The reference PyTorch implementation for Deform2Self is available on GitHub[1].

[1] https://github.com/daviddmc/Deform2Self.

## 3.2 Results

**Comparison with Other Approaches:** Table 1 shows the quantitative results on PINCAT and ACDC dataset. The proposed method achieves comparable or even better performance than other methods. Specifically, D2S has similar results to VBM4D in terms of Gaussian noise, and outperforms VBM4D for other noise models, especially for high noise levels. Besides, D2S outperforms S2S consistently, indicating that information from other time frames can largely boost performance of denoising models.

Figure 2 and 3 show example slices under Gaussian and Rician noise with $\sigma = 15\%$ and the denoised results using different methods. The D2S method has not only better statistical but also better perceptual results compared with other methods. Single-image methods such as BM3D and S2S, only use single noisy image, and therefore have not enough information to recover details that are corrupted by noise, resulting in blurred estimation. The DIP method suffers from significant structural artifacts in high noise levels. Though retrieving some details from adjacent frames, VBM4D also brings subtle artifacts to the denoised images. D2S is able to recover more detail structures with higher image quality.

**Table 1.** Quantitative results on PINCAT dataset for Gaussian and Poisson noise at different noise levels. The best results are indicated in red.

| PINCAT | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Gaussian | | | | | | Poisson | | | | | |
| | $\sigma = 15\%$ | | $\sigma = 20\%$ | | $\sigma = 25\%$ | | $P = 40$ | | $P = 20$ | | $P = 10$ | |
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Noisy | 16.55 | 0.300 | 14.05 | 0.208 | 12.11 | 0.151 | 22.42 | 0.603 | 19.40 | 0.472 | 16.19 | 0.346 |
| BM3D | 29.97 | 0.918 | 27.98 | 0.881 | 26.38 | 0.843 | 32.56 | 0.954 | 30.38 | 0.930 | 27.63 | 0.890 |
| VBM4D | 31.36 | 0.936 | 29.65 | 0.913 | 28.28 | 0.886 | 32.35 | 0.953 | 29.92 | 0.930 | 27.65 | 0.899 |
| DIP | 28.28 | 0.879 | 26.85 | 0.837 | 24.63 | 0.759 | 31.96 | 0.949 | 30.99 | 0.935 | 26.54 | 0.868 |
| S2S | 30.27 | 0.928 | 28.04 | 0.900 | 27.68 | 0.883 | 33.05 | 0.962 | 31.25 | 0.951 | 30.55 | 0.939 |
| D2S | 31.77 | 0.946 | 30.14 | 0.919 | 29.10 | 0.891 | 35.13 | 0.978 | 33.74 | 0.969 | 31.67 | 0.951 |
| ACDC | | | | | | | | | | | | |
| Method | Gaussian | | | | | | Rician | | | | | |
| | $\sigma = 5\%$ | | $\sigma = 10\%$ | | $\sigma = 15\%$ | | $\sigma = 5\%$ | | $\sigma = 10\%$ | | $\sigma = 15\%$ | |
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Noisy | 26.02 | 0.769 | 20.00 | 0.518 | 16.48 | 0.369 | 25.70 | 0.742 | 19.66 | 0.513 | 16.07 | 0.368 |
| BM3D | 32.32 | 0.953 | 28.54 | 0.905 | 26.45 | 0.860 | 29.58 | 0.874 | 23.69 | 0.777 | 19.78 | 0.689 |
| VBM4D | 32.54 | 0.957 | 28.96 | 0.911 | 26.88 | 0.863 | 29.79 | 0.879 | 23.94 | 0.791 | 19.93 | 0.707 |
| DIP | 26.95 | 0.875 | 25.55 | 0.815 | 23.48 | 0.718 | 26.10 | 0.811 | 22.76 | 0.736 | 19.10 | 0.629 |
| S2S | 30.41 | 0.942 | 28.45 | 0.912 | 26.90 | 0.880 | 28.28 | 0.861 | 23.51 | 0.784 | 19.73 | 0.709 |
| D2S | 32.16 | 0.960 | 30.26 | 0.936 | 28.22 | 0.887 | 29.37 | 0.879 | 24.25 | 0.812 | 20.20 | 0.743 |

**Ablation Study:** We perform ablation studies to evaluate different components in our framework. We evaluate models that ablate the single-image denoising module and STN module respectively.

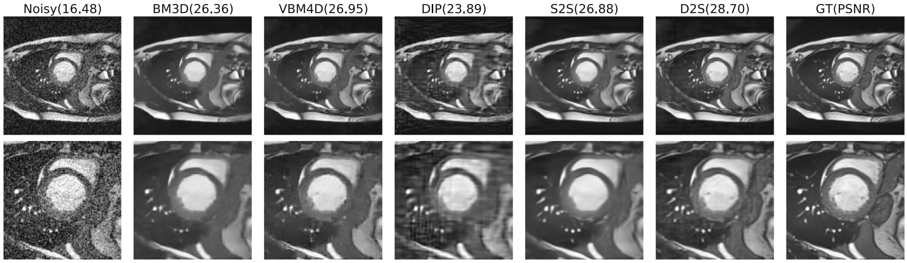

**Fig. 2.** Visualization results of different methods for ACDC dataset under Gaussian noise with $\sigma = 15\%$.

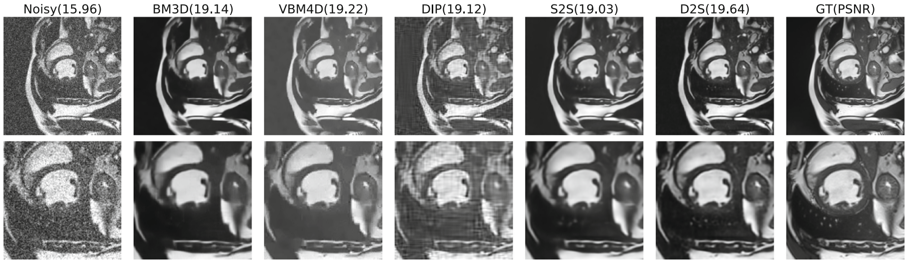

**Fig. 3.** Visualization results of different methods for ACDC dataset under Rician noise with $\sigma = 15\%$.

The ablated and full models are evaluated on the ACDC dataset with Gaussian and Rician noise ($\sigma = 10\%$). To investigate how single-image denoising improve registration accuracy and how image registration help match image content for the following multi-image denoising, we compute evaluation metrics PSNR and SSIM on the region of interest (ROI) that involves cardiac motion. The ROI masks include left ventricle, myocardium and right ventricle, which are annotated in the ACDC dataset [3].

Results of ablation studies (Table 2) indicate that the single-image denoising module improves the performance of the D2S model. It improves the accuracy of motion estimation in the registration network, and provides a good initialization for multi-image denoising. Besides, the registration module also makes contribution to the performance of our model. It registers images at different time frames to the same template so that the multi-channel input for the following denoising network is spatially matched, making the denoising task easier.

**Table 2.** Quantitative results of ablation studies.

| Method | Gaussian ($\sigma = 10\%$) | | Rician ($\sigma = 10\%$) | |
|---|---|---|---|---|
| | ROI-PSNR | ROI-SSIM | ROI-PSNR | ROI-SSIM |
| D2S | 28.01 | 0.894 | 27.55 | 0.889 |
| D2S w/o single-image denoising | 27.81 | 0.888 | 27.34 | 0.884 |
| D2S w/o image registration | 27.68 | 0.887 | 27.24 | 0.883 |

## 4 Conclusions

In this work, we proposed Deformed2Self, a self-supervised deep learning method for dynamic imaging denoising, which explores the similarity of image content at different time frames by estimating the motion during imaging and improve image quality with sequential single- and multi-image denoising networks. In addition, the proposed method only relies on the target noisy image with a small number of observations at other time frames and has no prerequisite on a large training dataset, making it more practical for applications with scarce data. Experiments on a variety of noise settings show that our method has comparable or even better performance than other state-of-the-art unsupervised or self-supervised denoising methods.

## References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
2. Batson, J., Royer, L.: Noise2Self: blind denoising by self-supervision. arXiv preprint arXiv:1901.11365 (2019)
3. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
4. Castillo, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. **54**(7), 1849 (2009)
5. Chen, H., et al.: Low-dose CT denoising with convolutional neural network. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 143–146. IEEE (2017)
6. Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3155–3164 (2018)
7. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
8. Gudbjartsson, H., Patz, S.: The Rician distribution of noisy MRI data. Magn. Reson. Med. **34**(6), 910–914 (1995)

9. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. arXiv preprint arXiv:1506.02025 (2015)
10. Jia, X., Liu, S., Feng, X., Zhang, L.: FOCNet: a fractional optimal control network for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6054–6063 (2019)
11. Johansson, A., Balter, J.M., Cao, Y.: Abdominal DCE-MRI reconstruction with deformable motion correction for liver perfusion quantification. Med. Phys. **45**(10), 4529–4540 (2018)
12. Kale, S.C., Chen, X.J., Henkelman, R.M.: Trading off SNR and resolution in MR images. NMR Biomed. Int. J. Devoted Dev. Appl. Magn. Reson. Vivo **22**(5), 488–494 (2009)
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2017)
14. Krull, A., Buchholz, T.O., Jug, F.: Noise2Void-learning denoising from single noisy images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2129–2137 (2019)
15. Lefkimmiatis, S.: Universal denoising networks: a novel CNN architecture for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3204–3213 (2018)
16. Lehtinen, J., et al.: Noise2Noise: learning image restoration without clean data. arXiv preprint arXiv:1803.04189 (2018)
17. Lingala, S.G., Hu, Y., DiBella, E., Jacob, M.: Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR. IEEE Trans. Med. Imaging **30**(5), 1042–1054 (2011)
18. Lukas, S., Feger, S., Rief, M., Zimmermann, E., Dewey, M.: Noise reduction and motion elimination in low-dose 4D myocardial computed tomography perfusion (CTP): preliminary clinical evaluation of the ASTRA4D algorithm. Eur. Radiol. **29**(9), 4572–4582 (2019)
19. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. IEEE Trans. Image Process. **21**(9), 3952–3966 (2012)
20. Malayeri, A.A., Johnson, W.C., Macedo, R., Bathon, J., Lima, J.A., Bluemke, D.A.: Cardiac cine MRI: quantification of the relationship between fast gradient echo and steady-state free precession for determination of myocardial mass and volumes. J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med. **28**(1), 60–66 (2008)
21. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2Self with dropout: learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1890–1898 (2020)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Schirrmacher, F., et al.: Temporal and volumetric denoising via quantile sparse image prior. Med. Image Anal. **48**, 131–146 (2018)
24. Sharif, B., Bresler, Y.: Adaptive real-time cardiac MRI using paradise: validation by the physiologically improved NCAT phantom. In: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1020–1023. IEEE (2007)
25. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)

26. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
27. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans. Med. Imaging **36**(12), 2536–2545 (2017)
28. Xie, D., et al.: Denoising arterial spin labeling perfusion MRI with deep machine learning. Magn. Reson. Imaging **68**, 95–105 (2020)
29. Xu, J., Gong, E., Ouyang, J., Pauly, J., Zaharchuk, G., Han, S.: Ultra-low-dose 18F-FDG brain PET/MR denoising using deep learning and multi-contrast information. In: Medical Imaging 2020: Image Processing, vol. 11313, p. 113131. International Society for Optics and Photonics (2020)
30. Xu, J., Gong, E., Pauly, J., Zaharchuk, G.: 200x low-dose pet reconstruction using deep learning. arXiv preprint arXiv:1712.04119 (2017)
31. You, X., Cao, N., Lu, H., Mao, M., Wanga, W.: Denoising of MR images with Rician noise using a wider neural network and noise range division. Magn. Reson. Imaging **64**, 154–159 (2019)
32. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)