Lesion focused super-resolution

Zhu, Jin, Yang, Guang, Lio, Pietro

Jin Zhu, Guang Yang, Pietro Lio, "Lesion focused super-resolution," Proc. SPIE 10949, Medical Imaging 2019: Image Processing, 109491L (15 March 2019); doi: 10.1117/12.2512576

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Lesion focused super-resolution

Jin Zhu[a], Guang Yang[b,c], and Pietro Lio[a]

[a] Department of Computer Science and Technology, University of Cambridge, Cambridge, CB3 0FD, UK
[b] Cardiovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK
[c] National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK

ABSTRACT

Super-resolution (SR) for image enhancement has great importance in medical image applications. Broadly speaking, there are two types of SR, one requires multiple low resolution (LR) images from different views of the same object to be reconstructed to the high resolution (HR) output, and the other one relies on the learning from a large amount of training datasets, i.e., LR-HR pairs. In real clinical environment, acquiring images from multi-views is expensive and sometimes infeasible. In this paper, we present a novel Generative Adversarial Networks (GAN) based learning framework to achieve SR from its LR version. By performing simulation based studies on the Multimodal Brain Tumor Segmentation Challenge (BraTS) datasets, we demonstrate the efficacy of our method in application of brain tumor MRI enhancement. Compared to bilinear interpolation and other state-of-the-art SR methods, our model is lesion focused, which has not only resulted in better perceptual image quality without blurring, but also been more efficient and directly benefit for the following clinical tasks, e.g., lesion detection and abnormality enhancement. Therefore, we can envisage the application of our SR method to boost image spatial resolution while maintaining crucial diagnostic information for further clinical tasks.

Keywords: Super-resolution, lesion detection, medical image analysis, image processing

1. INTRODUCTION

Images with high resolution (HR) are greatly in demand for many real applications.[1] However, the resolution and quality of the images are normally limited by the imaging hardware. For medical images, which provide useful and crucial details of the anatomical and physiological information for the patients, are very desirable with HR. In addition to the possible restrictions of the imaging hardware, medical images are more susceptible by the health limitations (e.g., ionizing radiation dose of using X-ray) and acquisition time limitations (e.g., Specific Absorption Rate limits of using MRI). Moreover, movements due to patients fatigue and organs pulsation will further degrade image qualities and result in images with lower signal-to-noise ratio (SNR). Low resolution (LR) medical images with limited field of view and degraded image quality could reduce the visibility of vital pathological details and compromise the diagnostic accuracy and prognosis.[2]

Research studies have shown that image super-resolution (SR) provides an alternative and relatively cheaper solution to improve the perceptual quality of medical images in terms of the spatial resolution enhancement instead of hardware improvement. Compared to conventional image interpolation, SR methods can provide better HR outputs with higher SNR and less blurry effects. Broadly speaking, there are two different types of SR: (1) using multiple LR images acquired from different views of the same object to reconstruct the HR output, but acquiring multi-view images could be expensive and sometimes infeasible; (2) learning a particular SR model using LR-HR training pairs, and performing the inference on a new input LR image to yield the HR output.[1,3]

More recently, deep learning based SR methods have boosted the performance of the super-resolved HR images mainly owe to the development of the computing power and the available big data. For example, deep residual neural networks(ResNet)[4] base SR methods successfully achieved high Peak-Signal-to-Noise-Ratio (PSNR) and Structural SIMilarity (SSIM) on nature image SR tasks.[5–7] Furthermore, Generative Adversarial Network helped

Further author information: (Send correspondence to Jin Zhu: E-mail: jin.zhu@cl.cam.ac.uk, Guang Yang: E-mail: g.yang@imperial.ac.uk and Pietro Lio: E-mail: pietro.lio@cl.cam.ac.uk)

Medical Imaging 2019: Image Processing, edited by Elsa D. Angelini, Bennett A. Landman, Proc. of SPIE
Vol. 10949, 109491L · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2512576

to demonstrate more perceptually realistic SR results in SRGAN.[7] However, all these methods were developed for only natural images and there are still limited studies for medical image super-resolution tasks.

In this study, we developed a lesion focused SR (LFSR) method that leverage the merits of GAN based models to generate perceptually more realistic SR results and also avoid introducing non-existing features into the lesion area after SR. By performing simulation based studies on the Multimodal Brain Tumor Segmentation Challenge (BraTS) datasets, we demonstrate the efficacy of our SR method in application of spatial resolution enhancement of brain tumor MRI images to potentially maintain crucial diagnostic information for further clinical tasks.

2. METHODS

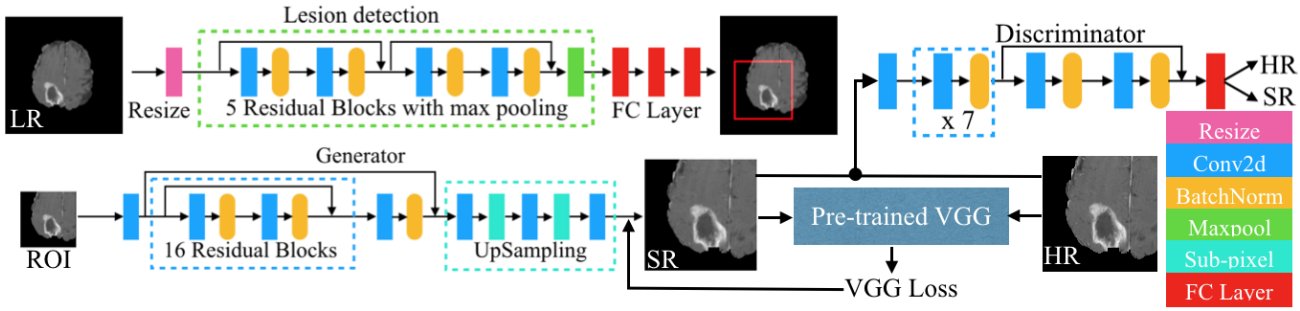

Figure 1. The schema of our proposed lesion focused super resolution (LFSR) neural network.

2.1 Lesion Focused SR

Our LFSR(Fig.1) includes a lesion detection neural network LD, a super resolution images generator G, a HR/SR images discriminator D, and a pre-trained 19 layers VGG.[8] The LD aims to detect the region of interest (ROI, e.g. brain tumors), $I_{\rm lr}$ and $I_{\rm hr}$, from whole size LR and HR images $I_{\rm LR}$ and $I_{\rm HR}$ before we applying the GAN:

$$I_{\rm lr,hr} = {\rm LD}(I_{\rm LR,HR}) \tag{1}$$

We propose a max pooling residual block and an input-scale free residual neural network LD. Compared to the residual blocks[4] and skip connection have been widely used, a max pooling layers is added after two residual blocks, which include two skip connections between four convolution and batch normalization layers. This can help accelerate the training process, and reduce the memory cost of the ROI detection task.

During the training of the GAN, G and D are playing a game: G aims to estimate as realistic as possible SR images, $I_{\rm SR}$, from $I_{\rm LR}$, and the discriminator aims to figure them out from the ground truth $I_{\rm HR}$. With the lesion detection, the training aims to solve:

$$(\hat{\theta_{\rm G}}, \hat{\theta_{\rm D}}) = \underset{\theta_{\rm G},\theta_{\rm D}}{\rm argmin} \sum l^{\rm G}(G_{\theta_{\rm G}}(I_{\rm lr}), I_{\rm hr}) + l^{\rm D}(D_{\theta_{\rm D}}(G_{\theta_{\rm G}}(I_{\rm lr}), I_{\rm hr}))) \tag{2}$$

where $\hat{\theta_{\rm G}}$ and $\hat{\theta_{\rm D}}$ are the trainable parameters, $l^{\rm G}$ and $l^{\rm D}$ are the loss functions for the G and D. In our proposed LFSR, we use a SR residual network (SRResNet) as the generator G , which includes 16 residual blocks, and following sub-pixel convolution layers. The discriminator D consists of 7 residual blocks of convolution layers and following fully connected layer, to distinguish the ground truth HR images from the generated SR ones. D and the pre-trained VGG are trained simultaneously with G to generate perceptually realistic image features.

2.2 Data Preprocessing and Training Settings

We have tested bilinear interpolation, SRResNet, SRGAN[7] and LFSR on the post-contrast T1-weighted (T1Gd) MRI scans from the BraTS 2018 datasets,[9,10] which have been randomly divided into training (9559 images) and validation (2368 images) datasets. All the images are normalized to zero-mean and unit-variance. We simulated

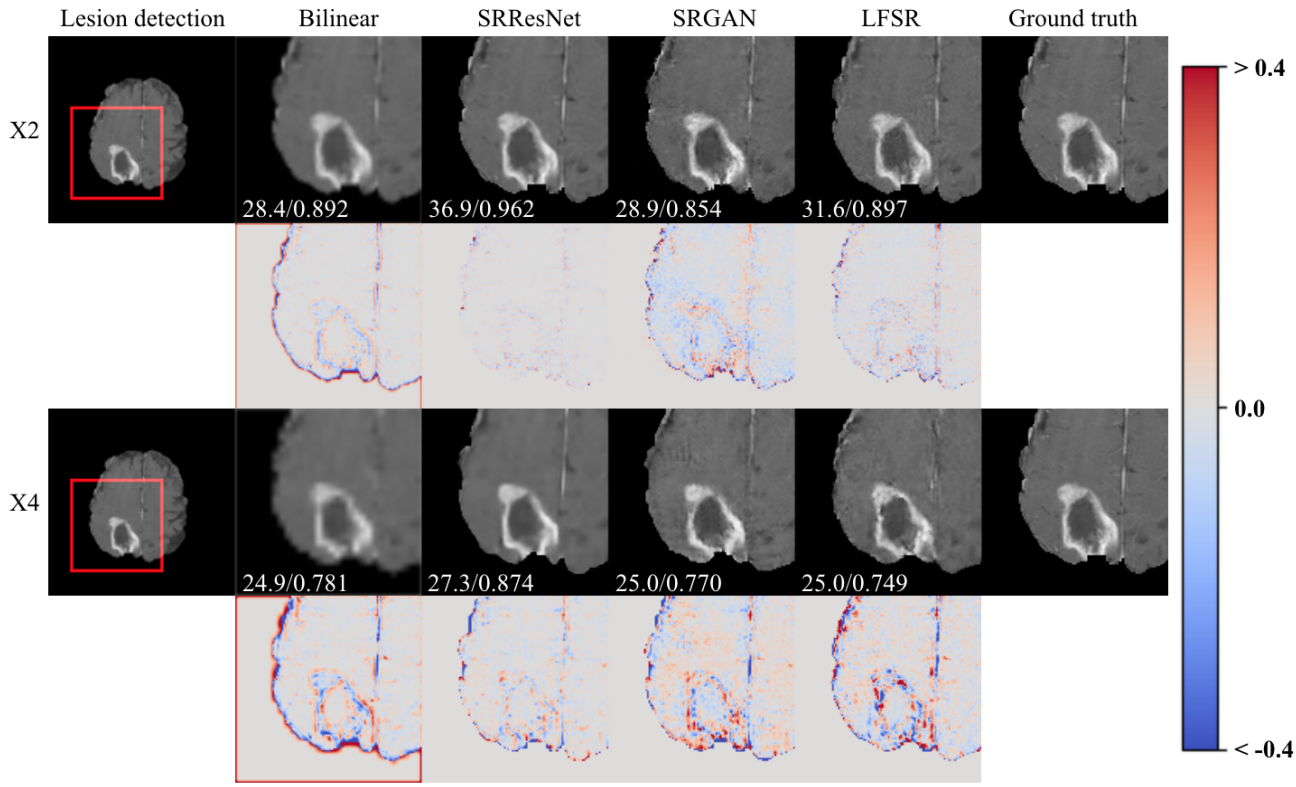

Figure 2. The ROIs of ground truth, the detected ROI with the predicted SR images(range: [-1, 1]), PSNR/SSIM of each result are also displayed.

the LR images by downsampling the HR ground truth and tested with additive white Gaussian noises (AWGN, $\sigma = 20, 40$) applied in the k-spaces.[11]

All the experiments were performed on a Linux workstation with NVIDIA TITAN Xp GPUs. All the models were implemented in Python, based on the TensorLayer[12] library, and were trained with Adam optimizer with the initial learning rate of 10^{-4}. The LD was trained independently for 100 epochs with L_2 loss. The SRResNet was trained for 350 epochs with the pixel-wise mean square error loss l_{MSE}. The generator in SRGAN and LFSR was initially trained with l_{MSE} for 50 epochs, then the GAN was trained with $l^{\text{G}} = l_{\text{MSE}} + l_{\text{VGG}} - l_{\text{GD}}$, and $l^{\text{D}} = 1 - l_{\text{DT}} - l_{\text{DF}}$, where l_{GD} was the percentage of incorrectly distinguished $I_{\text{SR,sr}}$, and l_{DT}, and l_{DF} were the percentages of correctly distinguished $I_{\text{HR,hr}}$ and $I_{\text{SR,sr}}$.

3. RESULTS AND DISCUSSIONS

3.1 Lesion Detection

Our LD has achieved high accuracies on both X2 and X4 downsampled images. In evaluation, we defined that if a tumor was 100% covered by the predicted ROI, it was a perfect detection, and if it was 95% covered, it was a acceptable detection. In the X2 case, 2218 images (93.7%) were perfect detections, and other 111 (98.4%) were acceptable detections. In the X4 case, 2109 images (89.1%) were perfect detections, and other 119 (94.1%) acceptable detections .

3.2 X2 and X4 SR

Here we show the X2 and X4 SR results in Fig. 2. Both bilinear interpolation and SRResNet have produced blurry SR results although SRResNet has achieved the highest PSNR and SSIM (Fig.3). SRGAN and our proposed LFSR have resulted in images with more realistic texture features compared to the ground truth. Compared

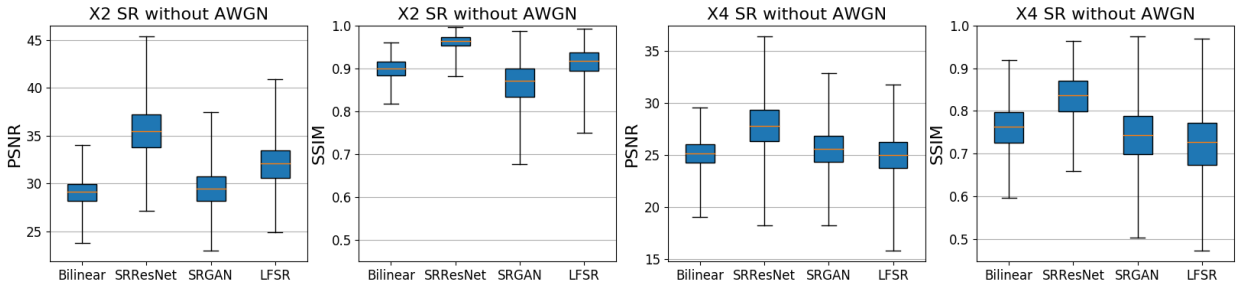

Figure 3. The PSNR and SSIM of the whole validation dataset (2368 images), in X2 and X4 SR tasks.

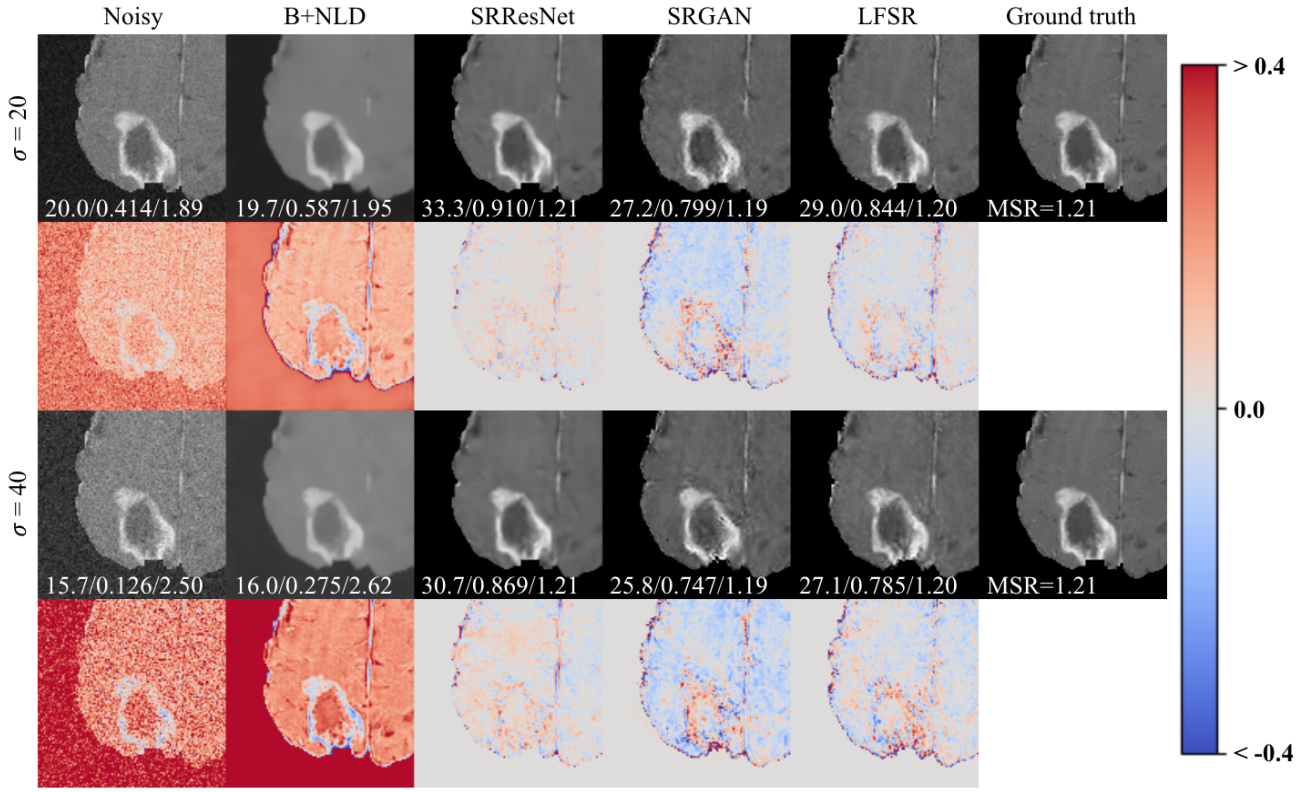

Figure 4. Two levels of AWGN ($\sigma = 20, 40$) have been applied in k-spaces in X2 super resolution experiments.

to SRGAN, LFSR has obtained higher (X2 cases) or equivalent (X4 cases) PSNR and SSIM. More importantly, LFSR has achieved significant reduction of the GPU memory cost; therefore, our LFSR can double the batch size, which has accelerated the training process to 266.8s/epoch for X2 and 194.8s/epoch for X4 (compared to SRGAN training time 649.8s/epoch for X2, and 370.8s/epoch for X4).

3.3 X2 SR with Additive Noise

We have also tested LFSR in X2 SR with additive Gaussian noise (Fig.4). The bilinear interpolation with non local means denoising[13] method (B+NLD) was tested to suppress the noise and provided a more fair comparison. All three deep learning methods have achieved higher PSNR and SSIM when noise is presented (Table 1, Fig.5). The l_{MSE} based SRResNet has still achieved the highest PSNR and SSIM. In contrast, both SRGAN and LFSR have been still able to generate more perceptually realistic textures from our qualitative studies. Furthermore,

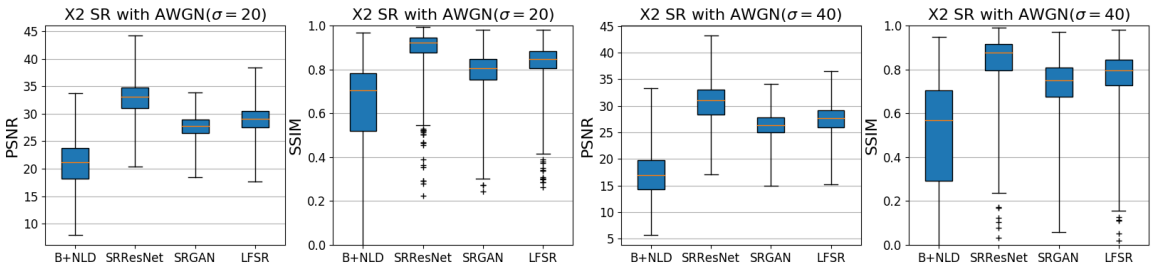

Figure 5. Two levels of AWGN ($\sigma = 20, 40$) have been applied in k-spaces in X2 super resolution experiments.

LFSR has achieved higher PSNR and SSIM than SRGAN for the noisy cases (Table 1), and more efficient training.

Table 1. PSNR and SSIM results for simulations with and without additive Gaussian noise (bold: better than SRGAN).

	PSNR				SSIM			
	X2	X4	X2($\sigma = 40$)	X2($\sigma = 40$)	X2	X4	X2($\sigma = 20$)	X2($\sigma = 40$)
B+NLD	29.1	25.2	20.7	17.1	0.900	0.761	0.623	0.483
SRResNet	35.6	27.9	32.8	30.6	0.962	0.832	0.895	0.840
SRGAN	29.6	25.7	27.6	26.2	0.865	0.741	0.789	0.731
LFSR	**32.1**	25.1	**29.0**	**27.4**	**0.914**	0.723	**0.832**	**0.772**

4. CONCLUSION

In summary, we have developed and validated a lesion focused SR (i.e., LFSR) method to super-resolve the tumor ROIs imaged by MRI. Compared to state-of-the-arts SR methods, our proposed LFSR method is more efficient and it can result in perceptually more realistic SR, which will maintain crucial image features for further clinical tasks and decisions. In the final camera ready version, we will include a more detailed description of our method and more comparison results.

ACKNOWLEDGMENTS

Jin Zhu's PhD research is funded by China Scholarship Council (grant No.201708060173). Guang Yang is funded by the British Heart Foundation Project Grant (Project Number: PG/16/78/32402).

REFERENCES

[1] Trinh, D. H., Luong, M., Dibos, F., Rocchisani, J.-M., Pham, C. D., and Nguyen, T. Q., "Novel example-based method for super-resolution and denoising of medical images.," *IEEE Transactions on Image Processing* **23**(4), 1882–1895 (2014).

[2] Yang, G., Ye, X., Slabaugh, G., Keegan, J., Mohiaddin, R., and Firmin, D., "Combined self-learning based single-image super-resolution and dual-tree complex wavelet transform denoising for medical images," in [*Medical Imaging 2016: Image Processing*], **9784**, 97840L, International Society for Optics and Photonics (2016).

[3] Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T., "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing* **21**(8), 3467–3478 (2012).

[4] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*CVPR*], 770–778 (2016).

[5] Tai, Y., Yang, J., and Liu, X., "Image super-resolution via deep recursive residual network," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], **1**(2), 5 (2017).

[6] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M., "Enhanced deep residual networks for single image super-resolution," in [*The IEEE conference on computer vision and pattern recognition (CVPR) workshops*], **1**(2), 4 (2017).

[7] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., et al., "Photo-realistic single image super-resolution using a generative adversarial network.," in [*CVPR*], **2**(3), 4 (2017).

[8] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

[9] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., "The multimodal brain tumor image segmentation benchmark (BRaTS)," *IEEE Transactions on Medical Imaging* **34**(10), 1993 (2015).

[10] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C., "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data* **4**, 170117 (2017).

[11] Bao, P. and Zhang, L., "Noise Reduction for Magnetic Resonance Images via Adaptive Multiscale Products Thresholding," *IEEE Transactions on Medical Imaging* **22**(9), 1089–1099 (2003).

[12] Dong, H., Supratak, A., Mai, L., Liu, F., Oehmichen, A., Yu, S., and Guo, Y., "TensorLayer: A Versatile Library for Efficient Deep Learning Development," *ACM Multimedia* (2017).

[13] Froment, J., "Parameter-free fast pixelwise non-local means denoising," *Image Processing On Line* **4**, 300–326 (2014).