Combined 3D super-resolution, de-noising and partial volume correction for percutaneous ablation

Mark A. Pinnock, Yipeng Hu, Steve Bandula and Dean C. Barratt

1. DESCRIPTION OF PURPOSE

Minimally invasive procedures are becoming more popular for the treatment of many conditions. Percutaneous cryoablation offers a lower rate of complications and more rapid recovery than radical nephrectomy for renal cell carcinoma.[1] This is achieved by using needles inserted in and around the tumour to cause cell death using extreme cold.[2] Interventional computed tomography (iCT) is a commonly used imaging modality for renal cryoablation, in part due to its excellent anatomical visualisation[3] as well as its competitive cost and availability in comparison to magnetic resonance (MR) imaging. However, no imaging system is free from artefacts or resolution constraints. To reduce the radiation exposure during iCT-guided procedures, the X-ray dose is decreased and the slice thickness is increased, resulting in increased noise and decreased spatial resolution in the z-direction.

Super-resolution (SR) is the application of post-processing techniques in an attempt to improve the above drawbacks in reducing radiation dose. Many sophisticated techniques have been developed to solve this inherently ill-posed problem, but are typically computationally expensive.[4,5] Convolutional neural networks (CNNs) have become widespread in SR applications since Dong et al.[6] proposed SRCNN and can achieve near real-time performance as demonstrated by Shi et al.,[7] an essential requirement for SR in interventional applications. The majority of previous work in CT has focused on increasing 2D in-plane resolution,[8–11] while three studies increased through-plane resolution (i.e. decreased slice thickness). Park et al.[12] trained a U-Net to learn a mapping between 15 mm and 3 mm slices, and later trained a network with residual connections to up-sample saggital images to increase the through-plane resolution.[13] Georgescu et al.[14] trained two sequential networks equipped with sub-pixel convolutional layers on patches to up-sample CT and MR images first in the xy-plane and then in the z-direction.

The aim of this work is to take the thick slice, noisy, low quality (LQ) images and convert them to thin slice, high quality (HQ) images. To achieve this, we make the following contributions: 1) We modify the 3D U-Net[15] with an up-sampling module that allows inputs and outputs of differing dimensions and trains on the entire image volume rather than patches; 2) As well as up-sampling in the z-direction, our proposed method also performs de-noising in the xy-plane; 3) Finding matched LQ-HQ pairs within iCT data is challenging owing to tissue deformation, respiratory motion, field of view (FOV) changes and movement within the scanner. We therefore train on simulated low quality (sLQ) data and show that this enables us to generalise to performing SR on real low quality (rLQ) images, tested on the scarce rLQ-HQ pairs. To our knowledge, this is the first application of SR on iCT images.

2. METHODS

The four-layer 3D U-Net used in this work uses skip layers to concatenate feature maps from the down-sampling blocks in the encoder with feature maps in their respective up-sampling blocks. Our input and output volumes are asymmetric, of size $512 \times 512 \times 3$ and $512 \times 512 \times 12$ respectively, requiring us to employ anisotropic kernels to down-sample and up-sample the images correctly. In addition, an up-sampling module up-samples the feature map skip connections with *tconv* kernels of size $1 \times 1 \times 3$ and stride $1 \times 1 \times 2$ before concatenation. The encoder consists of three down-sampling blocks, $dn(n) = [conv, conv, pool]$, with conv kernel of stride $1 \times 1 \times 1$, and pool kernel of size and stride $2 \times 2 \times 1$. The number of filters is $n = [8, 16, 32]$ and the encoder is followed by a *conv* layer of 64 filters. The decoder consists of 3 up-sampling blocks, $up(n) = [tconv, conv, conv]$ with $n = [32, 16, 8]$. The first block, $up(32)$, uses *tconv* of stride $2 \times 2 \times 1$, while $up(16)$ and $up(8)$ use *tconv* of stride $2 \times 2 \times 2$. All *conv* layers use filters of size $3 \times 3 \times 3$ and the ReLU activation function, aside from the final convolution layer which employed filters of size $1 \times 1 \times 1$ and a linear activation. Batch normalisation was not employed as in Ref 16.

The sLQ images were simulated in MATLAB (2018b) by blurring the HQ image volume in the z-direction and then averaging each sub-volume of four images to create the sLQ images. Thus, a volume of 12 thin slice images becomes a volume of 4 thick slice images. Noise was added in the projection domain after applying the Radon transform, and then the images were converted back to the image domain by filtered backprojection. These then formed the training examples, with mean squared error (MSE) used as a loss function.

The purpose of the following experiments was to determine: a) how well the proposed method can perform z-direction SR; b) how well it can perform de-noising; and c) whether it can be trained on simulated sLQ data and then generalise to rLQ iCT image volumes. The following sections describe the techniques to be used to test these hypotheses.

The retrospective data were fully anonymised after approval from the local clinical governance committee and consisted of images from iCT renal cryoablation procedures performed by the interventional oncology service at University College London Hospital. The training dataset consisted of 581 sLQ-HQ training volumes from 11 subjects, while the test dataset consisted of 35 rLQ and HQ volumes from 9 subjects. These were paired using 3D Slicer[17] to examine image volumes at the same coordinates – any mismatched volumes beyond subtle respiratory motion and other minor deformations were discarded. All images were normalised so that 0 and 1 corresponded to the maximum and minimum intensities of the CT scanner.

The network was trained for 500 epochs (taking approximately 48 hours) using Tensorflow 2.0[18] on a Nvidia P5000 16Gb graphics card. The minibatch size was fixed at 4 and the Adam optimiser was used for backpropagation ($\beta_1 = 0.9, \beta_2 = 0.999$). The optimal learning rate was found to be 0.001 after cross-validation.

The metrics used to evaluate image quality were peak signal-to-noise ratio (pSNR) and structural similarity index measure (SSIM). The proposed technique (UNet-SR) was compared with two interpolation techniques performed after blurring with a Gaussian filter: nearest neighbour (I-NN) and linear (I-LN). After preliminary experiments, the Gaussian filter that generated the highest pSNR and SSIM was found to have $\sigma_{xy} = 0.5$ and $\sigma_z = 1$ - this was used for all interpolations during testing. Because global measures such as these are not particularly sensitive to application-specific local regions of interest, we also employ the Sørenson-Dice coefficient (SDC) to compare needle localisation in the super-resolved image with the ground truth. Additional qualitative assessment is also reported in this work based on the visual comparison between the predicted HQ images and the paired real images.

3. RESULTS

The quantitative results are detailed in Table 1. The validation MSE, pSNR and SSIM (on sLQ data) for the proposed method are a substantial improvement over interpolation, demonstrating that the model performs very well on sLQ data. The improvement is less obvious during testing on the rLQ data, with only SSIM showing an incremental improvement. The effect on needle localisation is more pronounced, with a large improvement in SDC over that of interpolation.

The visual differences between the interpolation techniques and UNet-SR on rLQ data are apparent in the figures. The effect of optimal Gaussian blurring on the LQ noise in interpolation is minimal, while the super-resolved images from UNet-SR have been de-noised. Although there is some loss of detailed soft tissue texture when compared to ground truth, boundaries of high attenuation structures exhibit much better definition in the UNet-SR predictions than the comparison methods, which have a tendency to blur bony structures.

Regarding needle localisation, Figure 1 shows that the comparison methods have interpolated between the current rLQ slice and the next, causing the needle to appear incorrectly in the FOV. In contrast, despite the partial volume effect, UNet-SR has correctly sited the needle as shown by the arrows (although, like the interpolation methods, has not been able to predict the appearance of the haematoma around the needle insertion site in the successive ground truth image). Figure 2 shows the same correct needle localisation by UNet-SR; the network has correctly in-painted the needle shaft featured in the rLQ image but not present in the HQ image (yellow arrows). Figure 3 shows an example where UNet-SR has failed to correctly localise the needle. In the first row, we see that the network has erroneously extended the needle beyond its true position, while in the second row the needle has been incorrectly in-painted despite appearing in the ground truth image. Interestingly, the amount of lumbar vertebral disc (yellow arrows) visible in the UNet-SR images subtly changes from one slice to

	I-NN	I-LN	UNet
Val MSE	2.47e-5	2.35e-5	4.35e-6
Val pSNR	36.66	36.87	49.92
Val SSIM	0.9956	0.9959	0.9991
Test MSE	3.76e-5 [2.17e-5, 6.37e-5]	3.72e-5 [2.18e-5, 6.50e-5]	4.11e-5 [1.93e-5, 7.32e-5]
Test pSNR	40.30 [36.83, 41.74]	40.38 [36.81, 41.81]	39.43 [37.05, 41.81]
Test SSIM	0.9940 [0.9919, 0.9954]	0.9941 [0.9919, 0.9955]	0.9949 [0.9915, 0.9965]
Test Dice	0.2713 [0.2068, 0.3194]	0.2925 [0.2097, 0.3232]	0.4073 [0.1646, 0.5334]

Table 1. Validation metrics (averaged over validation folds) and test metrics for NN, linear interpolation and UNet. Test metrics given in the format: $q_{0.5}$ [$q_{0.1}$, $q_{0.9}$]
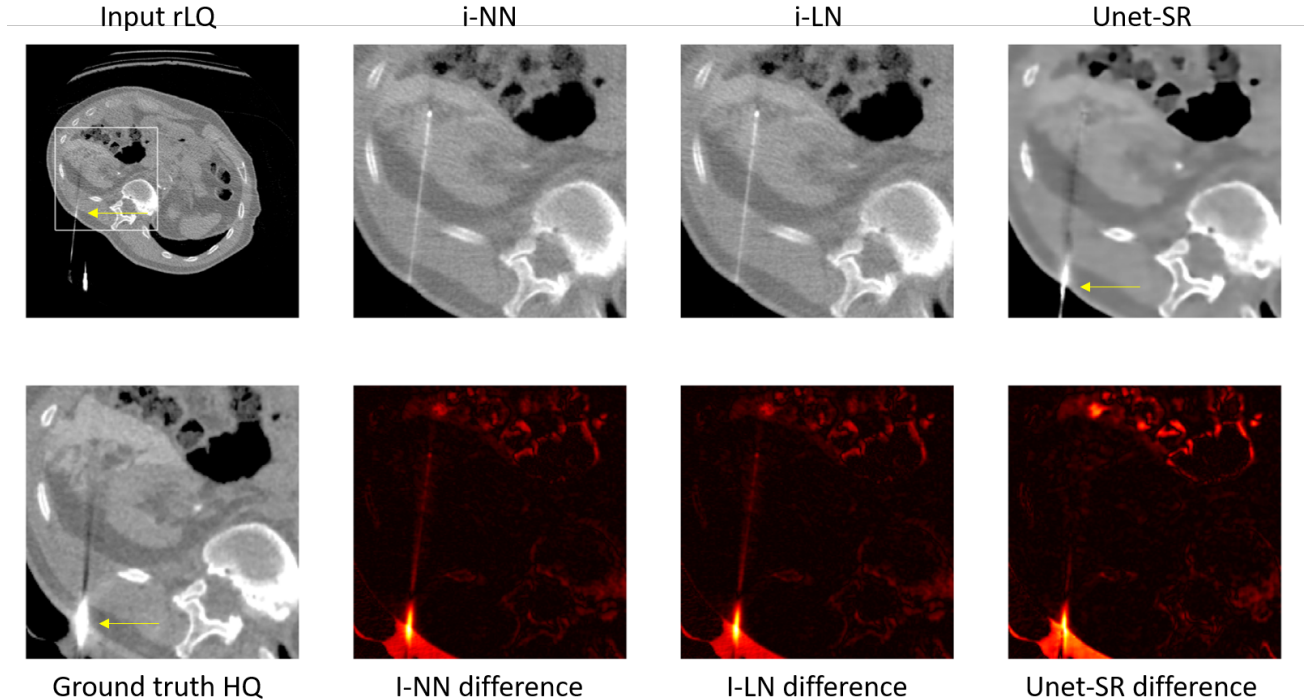

Figure 1. Input, ground truth and predicted images, along with absolute difference images. Arrows indicate correct needle prediction compared with interpolation.

the next as in the ground truth images, while the interpolation techniques do not show this level of detail - the neural network takes into account, and attempts to correct for, the partial volume effect.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We have trained a U-Net to up-sample iCT volumes in the z-direction, reducing the slice thickness and also performing de-noising in the process. In addition, we see that UNet-SR attempts to remove or add needles to the image by in-painting regions that may or may not contain them. While not out-performing interpolation-based techniques on real data to the same extent, we see that when validating on simulated data the network vastly outperforms interpolation, with MSEs an order of magnitude below that of the test data and a much higher pSNR and SSIM. This is likely explained by the real data being of a different distribution to the simulated volumes - further studies could optimise the data simulation step to more closely approximate the true data distribution. Additionally, the loss of detailed texture in the CNN output is a common problem in CNN-based SR techniques; the MSE loss has the tendency to average voxel intensities over the image, smoothing the resulting images. Indeed many studies have formulated priors to prevent such problems occurring and further work could explore this. We also highlight the difficulty of using global non-application-specific measures such as pSNR and SSIM

| Input rLQ | i-NN | i-LN | Unet-SR |

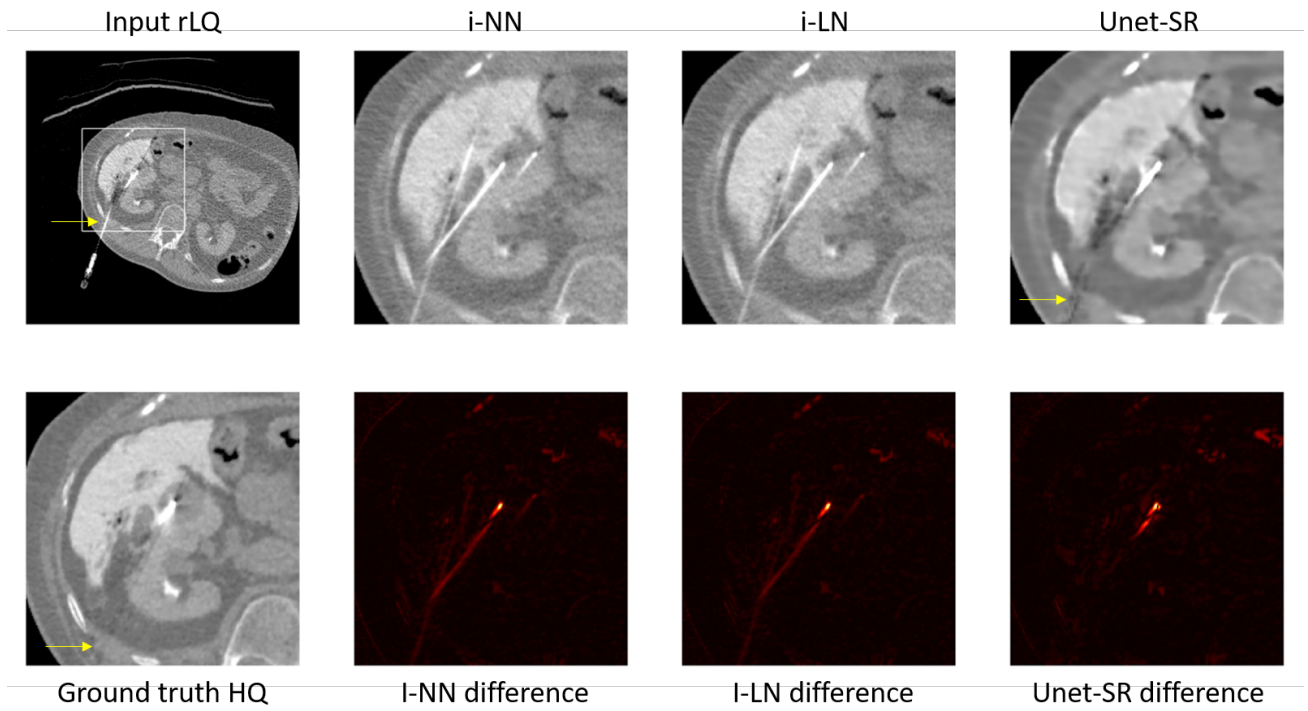| Ground truth HQ | I-NN difference | I-LN difference | Unet-SR difference |

Figure 2. Input, ground truth and predicted images, along with absolute difference images. Arrows indicate correct needle removal compared with interpolation.

| HQ slice 6 | i-NN slice 6 | i-LN slice 6 | Unet-SR slice 6 |

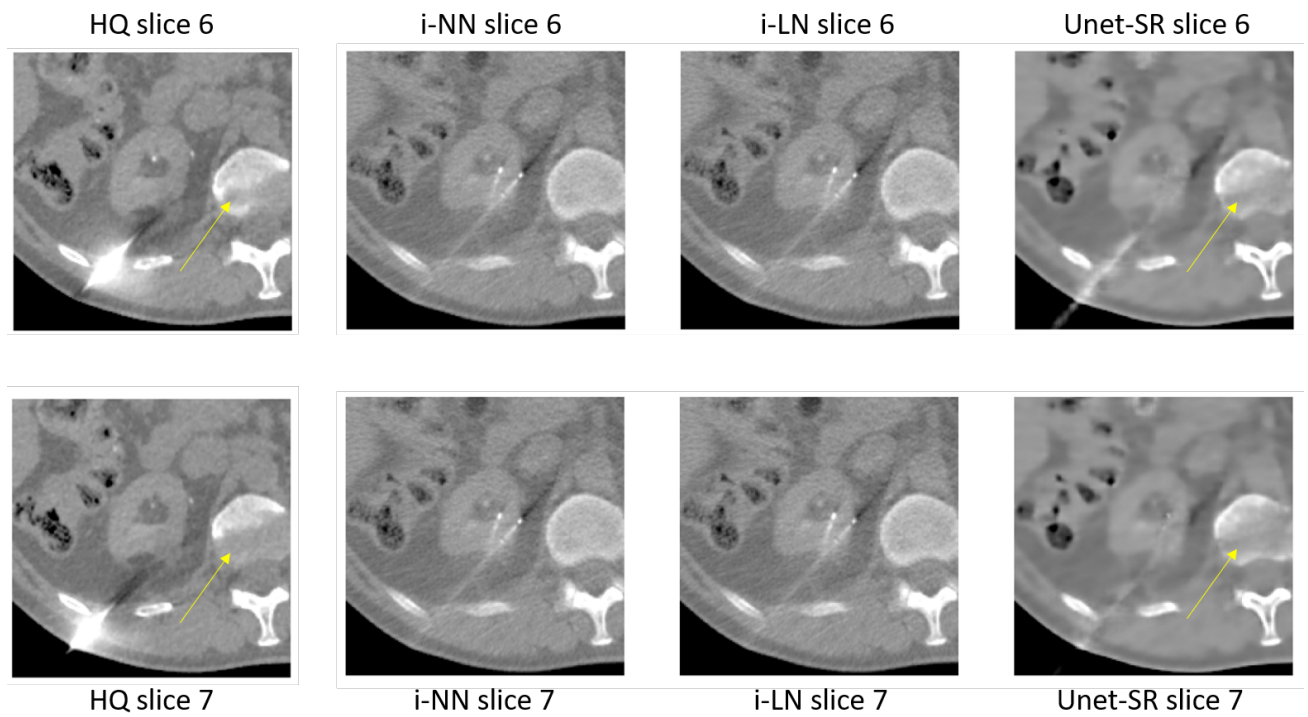| HQ slice 7 | i-NN slice 7 | i-LN slice 7 | Unet-SR slice 7 |

Figure 3. Top and bottom row: successive slices. Arrows show changes in lumbar vertebral disc between slice 6 and 7.

to assess the ability of a CNN to correct the partial volume effect for small but clinically relevant objects such as needles. In addition, the use of overlap-based metrics such as SDC are not well suited to long, thin objects. The development of application-specific metrics to assess or improve SR performance, for instance in needles or tissues of differing textures, is an open research topic.

5. CONCLUSION

In summary, we have shown that a CNN can be trained on simulated data to perform SR and de-noising on iCT data, outperforming common interpolation techniques and generating images with improved needle localisation, despite a different test data distribution to the training data. This was demonstrated in a challenging interventional radiology application using real clinical data.

6. OTHER SUBMISSIONS

This work has not been submitted for publication elsewhere.

REFERENCES

[1] Mues, A. C. and Landman, J., "Results of kidney tumor cryoablation: Renal function preservation and oncologic efficacy," *World Journal of Urology* **28**, 565–570 (Apr 2010).

[2] Uppot, R. N., Silverman, S. G., Zagoria, R. J., Tuncali, K., Childs, D. D., and Gervais, D. A., "Imaging-Guided Percutaneous Ablation of Renal Cell Carcinoma: A Primer of How We Do It," *American Journal of Roentgenology* **192**, 1558–1570 (Jun 2009).

[3] Permpongkosol, S., Nielsen, M. E., and Solomon, S. B., "Percutaneous renal cryoablation," *Urology* **68**, 19–25 (Jul 2006).

[4] Yang, J. and Huang, T., "Image super-resolution: Historical overview and future challenges," in [*Super-Resolution Imaging*], Milanfar P, ed., ch. 1, 3–35, CRC Press, Boca Raton, 1 ed. (2010).

[5] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q., "Deep Learning for Single Image Super-Resolution: A Brief Review," tech. rep. (2018).

[6] Dong, C., Loy, C. C., He, K., and Tang, X., "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(2), 295–307 (2016).

[7] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in [*Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], **2016-Decem**, 1874–1883 (Sep 2016).

[8] Umehara, K., Ota, J., and Ishida, T., "Application of Super-Resolution Convolutional Neural Network for Enhancing Image Resolution in Chest CT," *Journal of Digital Imaging* **31**, 441–450 (Aug 2018).

[9] Yu, H., Liu, D., Shi, H., Yu, H., Wang, Z., Wang, X., Cross, B., Bramlet, M., and Huang, T. S., "Computed Tomography Super-Resolution Using Convolutional Neural Networks," in [*IEEE International Conference on Image Processing (ICIP)*], 3944–3948 (2017).

[10] Li, M., Shen, S., Gao, W., Hsu, W., and Cong, J., "Computed Tomography Image Enhancement Using 3D Convolutional Neural Network," in [*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*], 291–299, Springer, Cham (Sep 2018).

[11] Liu, H., Xu, J., Wu, Y., Guo, Q., Ibragimov, B., and Xing, L., "Learning deconvolutional deep neural network for high resolution medical image reconstruction," *Information Sciences* **468**, 142–154 (Nov 2018).

[12] Park, J., Hwang, D., Kim, K. Y., Kang, S. K., Kim, Y. K., and Lee, J. S., "Computed tomography super-resolution using deep convolutional neural network," *Physics in Medicine & Biology* **63**, 145011 (Jul 2018).

[13] Park, S., Lee, S. M., Do, K. H., Lee, J. G., Bae, W., Park, H., Jung, K. H., and Seo, J. B., "Deep learning algorithm for reducing ct slice thickness: Effect on reproducibility of radiomic features in lung cancer," *Korean Journal of Radiology* **20**, 1431–1440 (Oct 2019).

[14] Georgescu, M. I., Ionescu, R. T., and Verga, N., "Convolutional Neural Networks with Intermediate Loss for 3D Super-Resolution of CT and MRI Scans," *IEEE Access* **8**, 49112–49124 (2020).

[15] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in [*Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*], **9901 LNCS**, 424–432, Springer Verlag (Jun 2016).

[16] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M., "Enhanced Deep Residual Networks for Single Image Super-Resolution," in [*Proceedings of the IEEE conference on computer vision and pattern recognition workshops*], (Jul 2017).

[17] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., and Kikinis, R., "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging* **30**, 1323–1341 (Nov 2012).

[18] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *arXiv preprint* (Mar 2016).