



Published in final edited form as:

*IEEE Trans Comput Imaging*. 2020 ; 6: 518–528. doi:10.1109/tci.2020.2964229.

## Super-Resolution PET Imaging Using Convolutional Neural Networks

Tzu-An Song, Samadrita Roy Chowdhury, Fan Yang, Joyita Dutta [Member, IEEE]

Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA, 01854 USA and co-affiliated with Massachusetts General Hospital, Boston, MA, 02114

### Abstract

Positron emission tomography (PET) suffers from severe resolution limitations which reduce its quantitative accuracy. In this paper, we present a super-resolution (SR) imaging technique for PET based on convolutional neural networks (CNNs). To facilitate the resolution recovery process, we incorporate high-resolution (HR) anatomical information based on magnetic resonance (MR) imaging. We introduce the spatial location information of the input image patches as additional CNN inputs to accommodate the spatially-variant nature of the blur kernels in PET. We compared the performance of shallow (3-layer) and very deep (20-layer) CNNs with various combinations of the following inputs: low-resolution (LR) PET, radial locations, axial locations, and HR MR. To validate the CNN architectures, we performed both realistic simulation studies using the BrainWeb digital phantom and clinical studies using neuroimaging datasets. For both simulation and clinical studies, the LR PET images were based on the Siemens HR+ scanner. Two different scenarios were examined in simulation: one where the target HR image is the ground-truth phantom image and another where the target HR image is based on the Siemens HRRT scanner — a high-resolution dedicated brain PET scanner. The latter scenario was also examined using clinical neuroimaging datasets. A number of factors affected relative performance of the different CNN designs examined, including network depth, target image quality, and the resemblance between the target and anatomical images. In general, however, all deep CNNs outperformed classical penalized deconvolution and partial volume correction techniques by large margins both qualitatively (e.g., edge and contrast recovery) and quantitatively (as indicated by three metrics: peak signal-to-noise-ratio, structural similarity index, and contrast-to-noise ratio).

### Keywords

super-resolution; CNN; deep learning; PET/MRI; multimodality imaging; partial volume correction

## I. Introduction

Positron emission tomography (PET) is a 3D medical imaging modality that allows *in vivo* quantitation of molecular targets. While oncology [1], [2] and neurology [3], [4] are perhaps the fields where PET is of the greatest relevance, its applications are expanding to many

---

J. Dutta is the corresponding author, dutta.joyita@mgh.harvard.edu.

other clinical domains [5], [6]. The quantitative capabilities of PET are confounded by a number of degrading factors, the most prominent of which are low signal-to-noise ratio (SNR) and intrinsically limited spatial resolution. While the former is largely driven by tracer dose and detector sensitivity, the latter is driven by a number of factors, including both physical and hardware-limited constraints and software issues. On the software front, resolution reductions are largely a product of smoothing regularizes and filters commonly used within or post-reconstruction for lowering the noise levels in the final images [7]. Together image blurring and tissue fractioning (due to spatial sampling for image digitization) lead to the so-called *partial volume effect* that is embodied by the spillover of estimated activity across different regions-of-interest (ROIs) [8].

Broadly, the existing computational approaches to address the resolution challenge encompass both within-reconstruction and post-reconstruction corrections. The former family includes methods that incorporate image-domain or sinogram-domain point spread functions (PSFs) in the PET image reconstruction framework [9]–[11] and/or smoothing penalties that preserve edges by incorporating anatomical information [12]–[19] or other transform-domain information [20]–[22]. The latter family of post-reconstruction filtering techniques includes both non-iterative partial volume correction methods [8], [23]–[27] and techniques that rely on an iterative deconvolution backbone [28]–[30] which is stabilized by different edge-guided or anatomically-guided penalty or prior functions [31], [32].

Unlike partial volume correction, strategies for which are often modality-specific, super-resolution (SR) imaging is a more general problem in image processing and computer vision. SR imaging refers to the task of converting a low-resolution (LR) image to a high-resolution (HR) one. The problem is inherently ill-posed as there are multiple HR images that may correspond to any given LR image. Classical approaches for SR imaging involve collating multiple LR images with subpixel shifts and applying motion-estimation techniques for combining them into an HR image frame [33]–[35]. More recently, many single-image SR (SISR) techniques have been proposed. Some SISR methods exploit self-similarities or recurring redundancies within the same image by searching for similar “patches” or sub-images within a given image [36]–[39]. Other SISR methods (the so-called “example-based” techniques) learn mapping functions from external LR and HR image pairs [40]–[44]. With the proliferation of deep learning techniques, LR-to-HR mapping functions based on deep neural networks have been demonstrated to yield state-of-the-art performance at SR tasks. One of the earliest deep SR approaches uses a 3-layer CNN architecture (commonly referred to in the literature as the SRCNN) [45]. Subsequently, a very deep SR (VDSR) CNN architecture [46] that had 20 layers and used residual learning [47] was shown to yield much-improved performance over the shallower SRCNN approach. Even more recently, SR performance has been further boosted by leveraging generative adversarial networks (GANs) [48], although these methods are somewhat limited by the challenges of GAN training, which remains notoriously difficult. While many recent papers have successfully utilized deep neural networks for PET image denoising and radiation dose reduction [49]–[51], the application of deep learning for SR PET imaging is a less-explored research territory [52], [53]. Unlike the denoising problem which aims to create a smoother image from a noisy one while ideally preserving edges, the SR problem aims to create a sharper image from a blurry one. Accordingly, SR requires distinct network design strategies

and data preparation than denoising. Our previous work on SR PET spans a wide gamut, including both (classical) penalized deconvolution based on joint entropy [32], [54] and deep learning approach based on the VDSR CNN [53]. This work is an extension of the latter effort.

In this paper, we design, implement, and validate several CNN architectures for SR PET imaging, including both shallow and very deep varieties. As a key adaptation of these models to PET imaging, we supplement the LR PET input image with its HR anatomical counterpart, e.g. a T1-weighted MR image. Unlike uniformly-blurred natural images, PET images are blurred in a spatially-variant manner [10], [55]. We, therefore, further adjust the network to accommodate spatial location details as inputs to assist the SR process. “Ground-truth” HR images are required for the training phase of all supervised learning methods. Though simulation studies are not constrained by this demand, it is a challenge for clinical studies where it is usually infeasible to obtain the “ground-truth” HR counterparts for LR PET scans of human subjects. To ensure clinical utility of this method, we extend it to training based on imperfect target images derived from a higher-resolution clinical scanner, thereby exploiting the SR framework to establish a mapping from an LR scanner’s image domain to an HR scanner’s image domain. In section II, we describe the underlying network architecture, the simulation data generation steps and the network training and validation procedures. In section III, we present simulation results comparing the performance of CNN-based SR with three well-studied reference approaches. A discussion of this method and our overall conclusions are presented in sections IV and V respectively.

## II. Methods

### A. Network Design

**1) Network Inputs:** One key contribution of this paper is the tailoring of the CNN inputs to address the needs of SR PET imaging. All CNNs implemented here have the LR PET as the main input. To further assist the resolution recovery process, additional inputs are incorporated as described below:

**a) Anatomical Inputs:** In generating the SR PET images, we seek to exploit the similarities between the PET and its high-resolution anatomical counterpart. Most clinical and preclinical PET scanners come equipped with anatomical imaging capabilities in the form of computed tomography (CT) or MR imaging to complement the functional information in PET with structural information. As illustrated in the schematic in Fig. 1, we employ CNNs with multi-channel inputs, that include LR PET and HR MR input channels.

**b) Spatial Inputs:** The spatial inputs are two 3D matrices of the same size as the image. One matrix stores the axial coordinates of each voxel, while the other stores the radial coordinates. In light of the cylindrical symmetry of PET scanners, these radial and axial inputs are sufficient for learning the spatially-variant structure of the blurring operator directly from the training data.

**c) Fusion:** For natural RGB images, where the three input channels typically have a high degree of structural similarity, usually the same set of kernels is effective for feature

extraction. In contrast, our CNN architecture exhibits a higher degree of input heterogeneity. Our initial experiments indicated the need for greater network width to accommodate a diverse set of features based on the very different input channels. We found an efficient solution to this by using separate kernels at the lower levels and fusing this information at higher levels as demonstrated in Fig. 1.

**2) Network Depth and Width:** A key finding of the VDSR paper was that the deeper the network, the better its performance. The paper showed that VDSR outperforms SRCNN by a great margin in terms of PSNR [46]. Here, we implement networks with different depths:

**a) Shallow CNN:** We designed and implemented a set of “shallow” networks loosely inspired by the 3-layer SRCNN [45]. This network only has three convolutional layers, each followed by a rectified linear unit (ReLU), except for the last output layer. Two key modifications were made to the 3-layer SRCNN to adapt it for SR PET imaging: 1) residual learning and 2) modified inputs as described in section II-A1. While, as a rule of thumb, the qualifier “deep” is applied to networks with three or more layers, we use the word “shallow” in this paper in a relative way.

**b) Very Deep CNN:** We designed and implemented a series of “very deep” CNNs based on the VDSR architecture in [46], but with a different input design and a fusion layer described in section II-A1. The very deep SR networks all have 20 convolutional layers, each followed by a ReLU, except for the last output layer.

For all the CNNs, the stride and padding for the convolution kernels were both set to 1. A  $3 \times 3$  kernel was adopted. All the convolutional layers had 64 filters except for the last layer, which had only 1 filter.

We implemented, validated, and compared both shallow (3-layer) and very deep (20-layer) CNN architectures with varying numbers of inputs. For the rest of the paper, we refer to these configurations as S1, S2, S3, S4, V1, V2, V3, and V4 as summarized in Table I.

**3) Network Implementation and Optimization:** All networks were trained using an  $L_1$  loss function. For an unknown SR PET image vectorized as  $\mathbf{x} \in \mathbb{R}^N$  and a target HR PET image  $\mathbf{x}_{\text{HR}} \in \mathbb{R}^N$ , where  $N$  is the number of voxels in the HR image domain, the  $L_1$  loss function is computed as:

$$\Phi_{\text{training}}(\mathbf{x} | \mathbf{x}_{\text{HR}}) = \|\mathbf{x} - \mathbf{x}_{\text{HR}}\|_1. \quad (1)$$

Although the  $L_2$  loss is the more frequently used variety of loss functions, recent literature [56] shows that the  $L_1$  loss achieves improved performance with respect to most image quality metrics such as the peak signal-to-noise ratio and the structural similarity index.

All networks were implemented on the PyTorch platform. Training was performed using GPU-based acceleration achieved by using an NVIDIA GTX 1080 Ti graphics card.

Minimization of the  $L_1$  loss function was performed using *Adam*, an algorithm for optimizing stochastic objective functions via adaptive estimates of lower-order moments [57].

## B. Simulation and Experimental Details

In the following, we describe two simulation studies using the BrainWeb digital phantom and the  $^{18}\text{F}$ -fluorodeoxyglucose ( $^{18}\text{F}$ -FDG) radiotracer and a clinical patient study also based on  $^{18}\text{F}$ -FDG. The inputs and targets for the studies are summarized in Table II. All LR PET images were based on the Siemens ECAT EXACT HR+ scanner. For the first simulation study, the HR PET images were the “ground-truth” images generated from segmented anatomical templates. For the second simulation study and the clinical study, the HR PET images were based on the Siemens HRRT scanner. Bicubic interpolation was used to resample all input and target images to the same voxel size of  $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$  with a  $256 \times 256 \times 207$  grid size. The LR and HR scanner properties are summarized in Table III [58].

### 1) Simulation Setup:

**a) *HR+ PSF Measurement:*** An experimental measurement of the true PSF was made by placing 0.5 mm diameter sources filled with  $^{18}\text{F}$ -FDG inside the HR+ scanner bore, which is 56.2 cm in diameter and 15.5 cm in length. The PSF images were reconstructed using ordered subsets expectation maximization (OSEM) with post-smoothing using a Gaussian filter. The PSFs were fitted with Gaussian kernels. We assumed radial and axial symmetry and calculated the PSFs at all other in-between locations as linear combinations of the PSFs measured at the nearest measurement locations. Interpolation weights for the experimental datasets were determined by means of bilinear interpolation over an irregular grid consisting of the quadrilaterals formed by the nearest radial and axial PSF sampling locations from a given point.

**b) *Target Image Generation for Study 1:*** Realistic simulations were performed using the 3D BrainWeb digital phantom (<http://brainweb.bic.mni.mcgill.ca/brainweb/>). 20 distinct atlases with 1 mm isotropic resolution were used to generate a set of “ground-truth” PET images. The atlases contained the following region labels: gray matter, white matter, blood pool, and cerebrospinal fluid. Static PET images were generated based on a  $\sim 1$  hour-long  $^{18}\text{F}$ -FDG scan as described in our earlier paper [32]. This “ground-truth” static PET is referred to as “true PET” for the rest of the paper. In Study 1, our purpose was to train the networks using the perfect “ground-truth” (clean target images). The target HR PET images are, therefore, the true PET images.

**c) *Target Image Generation for Study 2:*** In Study 2, we trained the networks using simulated HRRT PET images as our target HR images. The geometric model of the HRRT scanner was used to generate sinogram data. Poisson noise realizations were generated for the projected sinograms with a mean of  $10^8$  counts for a scan duration of 3640 s. The images were then reconstructed using the OSEM algorithm (6 iterations, 16 subsets). OSEM reconstruction results typically appear grainy due to noise. We, therefore, perform post-filtering with a 2.4 mm full width at half maximum (FWHM) 3D Gaussian filter. Since the

intrinsic resolution of the HRRT scanner is in the 2.3-3.4 mm range, this step improves image quality without any appreciable reduction in the resolution.

**d) Input Image Generation for Studies 1 and 2:** The geometric model of the HR+ scanner was used to generate sinogram data. Noisy data was generated using Poisson deviates of the projected sinograms, a noise model widely accepted in the PET imaging community [59]. The Poisson deviates were generated with a mean of  $10^8$  counts for the full scan duration of 3640 s. The data were then reconstructed using the OSEM algorithm (6 iterations, 16 subsets). The images were subsequently blurred using the measured, spatially-variant PSF to generate the LR PET images. In order to match HR PET image grid size, the LR PET images were interpolated into  $256 \times 256 \times 207$  from the HR+ output size of  $128 \times 128 \times 64$  using bicubic interpolation. T1-weighted MR images with 1 mm isotropic resolution derived directly from the BrainWeb database were used as HR MR inputs.

## 2) Experimental Setup:

**a) Clinical Data Source:** Clinical neuroimaging datasets for this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI, <http://adni.loni.usc.edu/>) database, a public repository containing images and clinical data from 2000+ human datasets. We selected 30 HRRT PET scans and the anatomical T1-weighted MPRAGE MR scans for clinical validation of our method. For consistency, all datasets were based on the ADNI1 protocol. 10 of the 30 subjects were from the cognitively normal (CN) category. The remaining 20 subjects had mild cognitive impairment (MCI). The full scan duration was 30 minutes ( $6 \times 5$ -minute frames). The OSEM algorithm (6 iterations, 16 subsets) was used for reconstruction.

**b) Target Image Generation for Study 3:** The OSEM-reconstructed HRRT PET images were post-filtered using a 2.4 mm FWHM 3D Gaussian filter to generate target HR PET images for the clinical study. This post-filtering step, as previously explained in the context of Study 2, reduces the noise without substantially reducing the image resolution of the HRRT images.

**c) Input Image Generation for Study 3:** The LR counterparts of the HRRT images were generated by applying the measured spatially-variant PSF of the HR+ scanner to the OSEM-reconstructed HRRT images. While not directly derived from the HR+ scanner, the use of a measured image-domain PSF ensures parity in terms of spatial resolution with true HR+ images. Rigidly co-registered T1-weighted MR images with  $1 \times 1 \times 1$  voxels were used as HR MR inputs. Cross-modality registration was performed using FSL (<https://fsl.fmrib.ox.ac.uk>) [60], [61].

## 3) Training and Validation:

**a) Cohort Sizes:** A cohort size of 20 human subjects was used for Studies 1 and 2. The size was dictated by the number of subjects in the BrainWeb database. For Study 3, the cohort comprised a total of 30 human subjects. The training and validation subsets each included a mix of MCI and CN subjects. For all the studies, training was performed using 15 subjects to predict a residual image that is an estimated difference between the input LR



PET and the ground truth HR PET. For Studies 1 and 2, the validation dataset consisted of 5 simulation subjects, while for Study 3, the validation dataset consisted of 15 clinical subjects.

**b) Input Dimensions:** We trained all the networks using multi-channel 2D transverse patches. For instance, when training S4 and V4, we sliced the four 3D input channels with combined dimensions of  $256 \times 256 \times 207 \times 4$  into 207 samples of size  $256 \times 256 \times 4$  (2D slices + 4 channels). The input intensities were normalized into the range  $[0, 1]$ . Data augmentation was achieved by randomly rotating the inputs by 1 to  $360^\circ$  and randomly cropping them to the size of  $96 \times 96 \times 4$ .

**c) Training Parameters:** Consistent selections of training parameters were used for all the CNN types and for all the studies. The learning rate was initially set to 0.0003 and decreased by  $5 \times 10^{-7}$  at every epoch. The batch size was 10. The networks were trained for 400 epochs.

### C. Reference Approaches

We compared the SR images generated by the networks S1, V1, S2, V2, S3, V3, S4, and V4 with the following approaches: deconvolution stabilized by a joint entropy (JE) penalty based on an anatomical MR image, deconvolution stabilized by the total variation (TV) penalty, and region-based voxel-wise (RBV) correction (a partial volume correction technique).

For an unknown SR PET image vectorized as  $\mathbf{x} \in \mathbb{R}^N$  and a scanner-reconstructed LR PET image  $\mathbf{b} \in \mathbb{R}^n$ , the least squares cost function for data fidelity is:

$$\Phi_{\text{PSF}}(\mathbf{x} | \mathbf{b}) = \frac{1}{2}(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}). \quad (2)$$

Here  $N$  and  $n$  are the respective numbers of voxels in the SR/HR and LR images and  $\mathbf{A} \in \mathbb{R}^{n \times N}$  represents a blurring and downsampling operation based on the measured spatially-variant PSF of the scanner [32]. The two deconvolution techniques used as reference in this paper seek to minimize a composite cost function,  $\Phi_{\text{PSF}} + \beta\Phi_{\text{reg}}$ , where  $\beta$  is a regularization parameter and  $\Phi_{\text{reg}}$  is a regularization penalty function. Given an HR MR image, denoted by a vector  $\mathbf{y} \in \mathbb{R}^N$ , the JE penalty is defined as [32]:

$$\Phi_{\text{reg-JE}}(\mathbf{x} | \mathbf{y}) = - \sum_{i=1}^M \sum_{j=1}^M \delta u \delta v p(u_i, v_j) \log p(u_i, v_j). \quad (3)$$

Here  $\mathbf{u} \in \mathbb{R}^M$  and  $\mathbf{v} \in \mathbb{R}^M$  are intensity histogram vectors based on the PET and MR images respectively and  $M$  is the number of intensity bins [32]. The TV penalty is defined as:

$$\Phi_{\text{reg-TV}}(\mathbf{x}) = (\|\Delta_1 \mathbf{x}\|_1 + \|\Delta_2 \mathbf{x}\|_1 + \|\Delta_3 \mathbf{x}\|_1), \quad (4)$$

where  $\nabla_k$  ( $k = 1, 2, \text{ or } 3$ ) are finite difference operators along the three cartesian coordinate directions [32].

The RBV technique [26] is a voxel-based extension of the popular geometric transfer matrix (GTM) method [8], a popular partial volume correction technique used in PET. It computes the corrected image  $\mathbf{x} \in \mathbb{R}^N$  from an upsampled blurry image of the same size,  $\tilde{\mathbf{b}} \in \mathbb{R}^N$  as follows:

$$\begin{aligned} \mathbf{x} &= \tilde{\mathbf{b}} \odot [\mathbf{s} \oslash (\mathbf{s} \otimes \mathbf{h})], \\ \mathbf{s} &= \sum_{i=1}^{N_{ROI}} g_i \mathbf{p}_i. \end{aligned} \quad (5)$$

Here  $\mathbf{h} \in \mathbb{R}^N$  is the approximate, spatially-invariant PSF,  $\odot$  and  $\oslash$  represent Hadamard (entrywise) multiplication and division respectively,  $\otimes$  represents 3D convolution,  $N_{ROI} = 2$  represents the two gray and white matter anatomical ROIs,  $\mathbf{p}_i$  is a binary mask for the  $i$ th ROI ( $p_{ij} = 1$  if voxel  $j$  belongs to ROI  $i$ , 0 otherwise), and  $g_i$  is the GTM-corrected intensity for the  $i$ th ROI [8].

#### D. Evaluation Metrics

The evaluation metrics used in this paper are defined below. In the following, the true and estimated images are denoted  $\mathbf{x}$  and  $\hat{\mathbf{x}}$  respectively. We use the notation  $\mu_x$  and  $\sigma_x$  respectively for the mean and standard deviation of  $\mathbf{x}$ .

**1) Peak Signal-to-Noise Ratio (PSNR):** The PSNR is the ratio of the maximum signal power to noise power and is defined as:

$$\text{PSNR}(\hat{\mathbf{x}}, \mathbf{x}) = 20 \log_{10} \left( \frac{\max(\hat{\mathbf{x}})}{\text{RMSE}(\hat{\mathbf{x}}, \mathbf{x})} \right), \quad (6)$$

where the root-mean-square error (RMSE) is defined as:

$$\text{RMSE}(\hat{\mathbf{x}}, \mathbf{x}) = \sqrt{\frac{1}{N} \sum_k (\hat{\mathbf{x}}_k - \mathbf{x}_k)^2}. \quad (7)$$

**2) Structural Similarity Index (SSIM):** The SSIM [62] is a well-accepted measure of perceived image quality and is defined as:

$$\text{SSIM}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}. \quad (8)$$

Here  $c_1$  and  $c_2$  are parameters stabilizing the division and  $\sigma_{x\hat{x}}$  is the covariance of  $\mathbf{x}$  and  $\hat{\mathbf{x}}$ .



**3) Contrast-to-Noise Ratio (CNR):** The CNR for a target ROI  $\mathcal{R}$  and a reference ROI  $\mathcal{R}_{\text{ref}}$  is defined as:

$$\text{CNR} = \frac{|\mu_{\hat{x}^{\mathcal{R}}} - \mu_{\hat{x}^{\mathcal{R}_{\text{ref}}}}|}{\sqrt{[(\sigma_{\hat{x}^{\mathcal{R}}})^2 + (\sigma_{\hat{x}^{\mathcal{R}_{\text{ref}}})^2]}}. \quad (9)$$

In this work, the gray matter ROI is treated as the target and the white matter ROI as the reference.

### III. Results

#### A. Study 1: Simulation Results

Fig. 2a showcases results from Study 1 — transverse slices from the HR MR, HR PET (same as the true PET in this case), LR PET, RBV PET, and the SR imaging results from the two deconvolution methods and eight CNNs described in Table I. The image frames are from one subject from the validation dataset. Magnified subimages in Fig. 2b highlight artifacts/inaccuracies indicated by purple arrows that are observed in all the techniques that lack anatomical guidance, namely TV, S1, V1, S3, and V3. Comparison of the subimage pairs (S1, S3) and (V1, V3) illustrates that, in the absence of anatomical information, spatial information greatly enhanced image quality. Comparison of the subimage pairs (S1, V1) and (S3, V3) also shows that the addition of more convolutional layers (increased network depth) is also very effective in the absence of anatomical information. Among the non-deep-learning methods, JE and RBV, which incorporate anatomical information, show better edge recovery than TV. The performance metrics (PSNR, SSIM, and gray-to-white CNR) for the different methods are tabulated in Table IV. The values of these metrics computed for the HR images (true PET images for Study 1) are also included in this table for reference. This table indicates that methods with anatomical information perform better than methods without this information. CNN-based approaches that incorporate anatomical information vastly exceed the performance of classical methods.

In terms of PSNR, the supervised CNNs outperform the classical approaches by a wide margin. S2, S4, V2, and V4, the networks that have anatomical guidance, show stronger performance in terms of PSNR, SSIM and CNR than S1, V1, S3, and V3, the networks which lack anatomical guidance. The PSNR, SSIM, and CNR values for S1 vs. S3 show that these metrics increase noticeably with the addition of spatial information for the shallow case.

#### B. Study 2: Simulation Results

Fig. 3a showcases results from Study 2 — transverse slices from the HR MR, true PET, HR PET, LR PET, RBV PET, and the SR imaging results from the two deconvolution methods and eight CNNs described in Table I. The displayed images are all from the same subject from the validation dataset. As with Study 1, the CNNs with MR-based anatomical inputs (S2, V2, S4, and V4) continue to produce the best results. However, since the target image is

now a corrupt image with diminished structural similarity with the MR image, the margin of gain from using anatomical information is now reduced. Magnified subimages in Fig. 3b highlight artifacts/inaccuracies indicated by purple arrows that are more preponderant in this study compared to Study 1. Interestingly, for this more challenging problem, the deeper networks (V2 and V4) have reduced background noise variations than their shallower counterparts (S2 and S4), as indicated by purple arrows in the latter. It should be noted that the JE, TV, and RBV images, which are unsupervised, are the same as those showcased in Fig. 2.

The PSNR, SSIM, and gray-to-white CNR of the different methods are tabulated in Table V. The values of these metrics computed for the HR images (HRRT PET images for Study 2) are also included in this table for reference. An additional goal for this study was to understand the variability in results that could be anticipated when imperfect HR images were used for training. We, therefore, computed two sets of PSNR, SSIM, and CNR measures: one with respect to the target HR PET and another with respect to the true PET. Our results show that there is an overall reduction in performance when the true PET is used as the reference. This is expected with the HR PET used for training deviates substantially from the true PET. But a key observation here is that the CNNs exhibit consistent relative levels of accuracy for the two reference images. As with Study 1, anatomically-guided networks showed better performance than the non-anatomically guided networks and the three classical methods. Also, interestingly, when the true PET is used as the reference, all CNN-based SR images show lower PSNR values than the HR image (HRRT in this case). This can be attributed to the denoising properties of CNNs that lead to SR results that are smoother than the noisy and imperfect target images used for training.

### C. Study 2: Network Depth and Width

To examine the impact of CNN depth and width, we extended Study 2 to compare three different network depths (3, 12 and 20 layers) and two different network widths (32 and 64 filters). More specifically, all the networks compared had either 32 or 64 filters in each layer, except for their output layers which had only 1 filter. For this study, all networks were trained using the 4-channel inputs (LR PET, HR MR, radial coordinates, and axial coordinates) to ensure a fair comparison. As shown in Table VI, the 20-layer network yielded the best performance in terms of PSNR and SSIM, while the 12-layer network outperformed the 3-layer network. Also, as expected, the networks with 64 filters led to consistently higher PSNR and SSIM than those with 32 filters. Fig. 4 shows the plots of mean squared error (MSE) against the epoch numbers. As demonstrated in this figure, the 20-layer network leads to the fastest convergence rate, followed the 12-layer and 3-layer networks respectively.

### D. Study 3: Experimental Results

Fig. 5a showcases results from Study 3 — the HR MR, HR PET, LR PET, RBV PET, and the SR imaging results from the two deconvolution methods and eight CNNs described in Table I. For this study, the deeper networks (V1, V2, V3, and V4) produced visually sharper images than their shallower counterparts (S1, S2, S3, and S4). This is clearly evident from the magnified subimages in Fig. 5b. This is consistent with our previous observation that, in

the absence of a strong contribution of the anatomical inputs, the extra layers lead to a stronger margin of improvement. That said, V4, which is deeper and uses anatomical information, led to the highest levels of gray matter contrast as highlighted by red arrows.

The PSNR, SSIM, and gray-to-white CNR of the different methods are tabulated in Table VII. The values of these metrics computed for the HR images (HRRT PET images for Study 3) are also included in this table for reference. As displayed in the table, V4 continues to exhibit the best performance in terms of PSNR, SSIM, and CNR. As in Studies 1 and 2, all CNN-based methods outperformed RBV, JE, and TV.

## IV. Discussion

Overall, our results indicate that CNNs vastly outperform penalized deconvolution and partial volume correction at the SR imaging task. Among different CNN architectures, relative performance depends on the problem at hand. Our simulation and clinical studies all agree that deep CNNs outperform shallow CNNs and that the additional channels contribute to the improvement of overall performance. Like network depth, network width also helped improve performance. These improvements, however, come at the expense of added computational cost. With 4-channel inputs, for example, the training times for 3-, 12-, and 20-layer networks based on 64 filters were 4.7, 7.8, and 12.2 hours respectively. Furthermore, based on our experiments, we concluded that the relative importance of anatomical and spatial input channels depends on the underlying structural similarity between the HR MR and the true PET.

One limitation of our simulation study is the limited size of the validation dataset, which consisted of only 5 subjects. This limitation was overcome in the clinical datasets, where the validation cohort size was increased to 15 subjects. We observed that the performance trends of the compared methods are similar for the experimental and simulation studies. This observation enhances the credibility of our simulation results, despite the small cohort size.

Another limitation of the presented SR approach is the lack of portability of the trained models to inputs with image characteristics (e.g., noise, blur, contrast) substantially different from those used for training. This is a shortcoming that is shared by most supervised learning models. The results presented here correspond to the HRRT-HR+ scanner pair and the  $^{18}\text{F}$ -FDG radiotracer. Additionally, our experimental study relied on data based on a single acquisition protocol. While the network can be easily retrained for alternative scanner pairs, tracer types, and acquisition protocols, the performance characterization may need to be repeated for these specific scenarios. The SR performance is expected to drop when the LR inputs are based on a much lower or higher tracer dose than the datasets used for training. Our future work will characterize the performance of these networks when input noise levels are varied. One remedial approach to address noise sensitivity is the use of transfer learning for easy retraining of the networks with a much smaller training dataset containing LR inputs with altered SNR. Another more sophisticated strategy is to use the CNN output as a prior in reconstruction. We have previously used this approach and produced promising results for denoising [63] and anticipate that it will work for deblurring and SR by extension.

The presence of axial and radial inputs makes the overall network well-equipped to handle 3D inputs as well as 2D slice stacks from all three coordinate directions in a given training dataset. In the current analysis, we used 2D transverse slices alone for training and validation. In this setting, the axial inputs are particularly critical for modeling axial variation of resolution. The radial slices offer some redundancy as the radial location information for each voxel could be derived from the PET or MR image slices directly. However, based on our experience, explicitly providing this information serves as additional feature guidance and is helpful in tackling spatially-variant blurring. Due to the computational cost, we resorted to 2D convolutions instead of 3D. 3D kernels are expected to perform better than 2D ones. However, with 3D kernels, the computational cost is expected to skyrocket. This could prove to be a major bottleneck toward implementing a very deep network (a key contribution of this paper).

In this paper, we exclusively focused on neuroimaging applications. However, the methodology is well generalizable to other parts of the body subject to the availability of appropriate training data. It should be noted that the utilization of anatomical images in conjunction with functional images required co-registration. To ensure robustness, a well-tested standardized registration tool was used for the task. Since the registration is intra-subject, rigid registration based on mutual information suffices for this application.

Another limitation of CNN-based SR (and perhaps the most significant one) is that it relies on supervised learning and, therefore, requires paired LR PET and HR PET images for training. This requirement is easy to address while training using simulated datasets. But paired LR and HR clinical scans are rare. To address this limitation, we are currently exploring self-supervised learning strategies based on adversarial training of generative adversarial networks, that circumvent the need for paired training inputs.

## V. Conclusion

We have designed, implemented, and validated a family of CNN-based SR PET imaging techniques. To facilitate the resolution recovery process, we incorporated both anatomical and spatial information. In the studies presented here, the anatomical information was provided as an HR MR image. So as to easily provide spatial information, we supplied patches containing the radial and axial coordinates of each voxel as additional input channels. This strategy is well-consistent with standard CNN multi-channel input formats and, therefore, convenient to implement. Both simulation and clinical studies showed that the CNNs greatly outperform penalized deconvolution both qualitatively (e.g., edge and contrast recovery) and quantitatively (as indicated by PSNR, SSIM, and CNR).

While PSNR and SSIM are metrics that are well-trusted in the image processing and computer vision communities, CNR is probably is a more clinically-meaningful performance metric. Our results confirm that the PSNR and SSIM improvements are also accompanied by improvements in CNR. We, therefore, conclude that SR PET imaging using very deep CNNs is clinically promising. As future work, we will develop new networks based on self-supervised learning that will enable us to circumvent the need for paired training datasets. We will also explore avenues to ensure the applicability of this method to

super-resolve inputs with noise levels different from those in the training data. Furthermore, as a strategy for ensuring more robust performance with clinical data, we plan to use a hybrid (simulation + clinical) dataset for CNN training and validation. From an applications perspective, we are interested in using this technique to super-resolve PET images of tau tangles, a neuropathological hallmark of Alzheimer's disease, with the goal of developing sensitive image-based biomarkers for tau [38].

## Acknowledgment

This work was supported in part by the NIH grant K01AG050711. The authors would like to thank Dr. Kyungsang Kim for his inputs on the simulation studies.

## References

- [1]. Farwell MD, Pryma DA, and Mankoff DA, "PET/CT imaging in cancer: current applications and future directions," *Cancer*, vol. 120, no. 22, pp. 3433–3445, 11 2014. [PubMed: 24947987]
- [2]. Sotoudeh H, Sharma A, Fowler KJ, McConathy J, and Dehdashti F, "Clinical application of PET/MRI in oncology," *J. Magn. Reson. Imaging*, vol. 44, no. 2, pp. 265–276, 8 2016. [PubMed: 27007987]
- [3]. Salmon E, Bernard Ir C, and Hustinx R, "Pitfalls and Limitations of PET/CT in Brain Imaging," *Semin. Nucl. Med.*, vol. 45, no. 6, pp. 541–551, 11 2015. [PubMed: 26522395]
- [4]. Catana C, Drzezga A, Heiss WD, and Rosen BR, "PET/MRI for neurologic applications," *J. Nucl. Med.*, vol. 53, no. 12, pp. 1916–1925, 12 2012. [PubMed: 23143086]
- [5]. Salata BM and Singh P, "Role of Cardiac PET in Clinical Practice," *Curr. Treat. Options Cardiovasc. Med.*, vol. 19, no. 12, p. 93, 11 2017. [PubMed: 29119262]
- [6]. Hess S, Alavi A, and Basu S, "PET-Based Personalized Management of Infectious and Inflammatory Disorders," *PET Clin.*, vol. 11, no. 3, pp. 351–361, 7 2016. [PubMed: 27321037]
- [7]. Qi J and Leahy RM, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imaging*, vol. 19, no. 5, pp. 493–506, 5 2000. [PubMed: 11021692]
- [8]. Rousset OG, Ma Y, and Evans AC, "Correction for partial volume effects in PET: principle and validation," *J. Nucl. Med.*, vol. 39, no. 5, pp. 904–911, 5 1998. [PubMed: 9591599]
- [9]. Reader AJ, Julyan PJ, Williams H, Hastings DL, and Zweit J, "EM algorithm system modeling by image-space techniques for PET reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 5, pp. 1392–1397, 10 2003.
- [10]. Alessio AM, Kinahan PE, and Lewellen TK, "Modeling and incorporation of system response functions in 3-D whole body PET," *IEEE Trans. Med. Imaging*, vol. 25, no. 7, pp. 828–837, 7 2006. [PubMed: 16827484]
- [11]. Panin VY, Kehren F, Michel C, and Casey M, "Fully 3-D PET reconstruction with system matrix derived from point source measurements," *IEEE Trans. Med. Imaging*, vol. 25, no. 7, pp. 907–921, 7 2006. [PubMed: 16827491]
- [12]. Leahy R and Yan X, "Incorporation of anatomical MR data for improved functional imaging with PET," in *Inf. Process. Med. Imaging*, vol. 511 Springer, 1991, pp. 105–120.
- [13]. Comtat C, Kinahan PE, Fessler JA, Beyer T, Townsend DW, Defrise M, and Michel C, "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels," *Phys. Med. Biol.*, vol. 47, no. 1, pp. 1–20, 1 2002. [PubMed: 11814220]
- [14]. Bowsher JE, Yuan H, Hedlund LW, Turkington TG, Akabani G, Badea A, Kurylo WC, Wheeler CT, Cofer GP, Dewhurst MW., et al. "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4 IEEE, 2004, pp. 2488–2492.
- [15]. Baete K, Nuyts J, Van Paesschen W, Suetens P, and Dupont P, "Anatomical-based FDG-PET reconstruction for the detection of hypometabolic regions in epilepsy," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 510–519, 4 2004. [PubMed: 15084076]

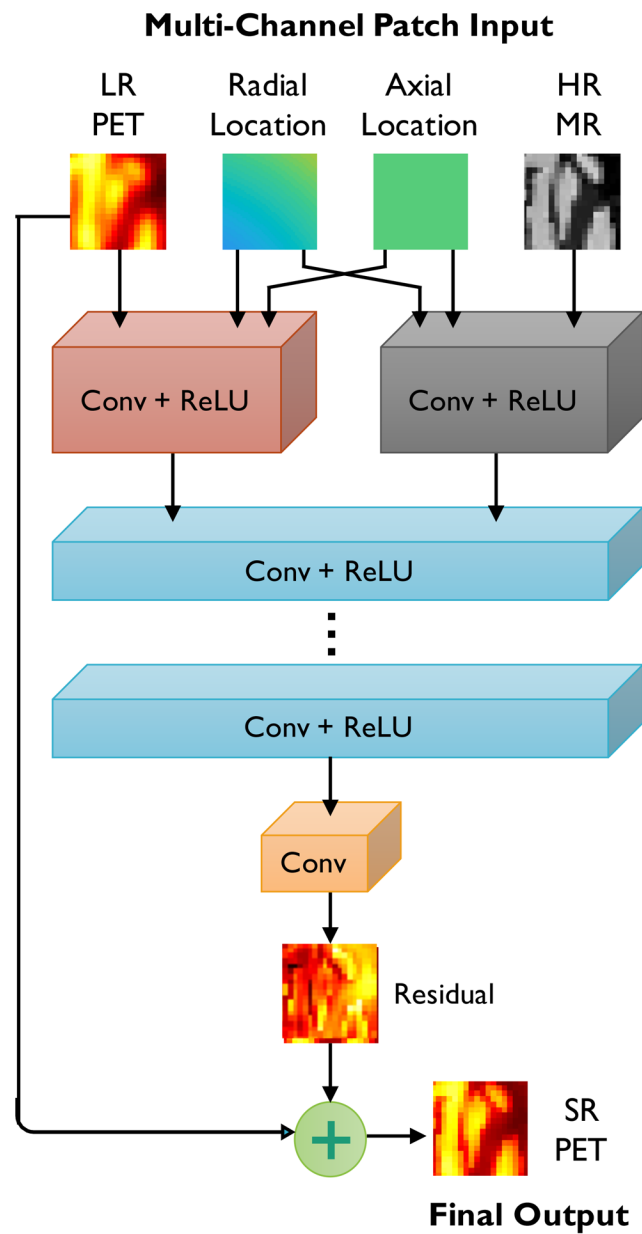
- [16]. Bataille F, Comtat C, Jan S, Sureau F, and Trebossen R, “Brain PET partial-volume compensation using blurred anatomical labels,” *IEEE Trans. Nucl. Sci.*, vol. 54, no. 5, pp. 1606–1615, 4 2007.
- [17]. Pedemonte S, Bousse A, Hutton BF, Arridge S, and Ourselin S, “4-D generative model for PET/MRI reconstruction,” *Med. Image Comput. Comput. Assist. Interv.*, vol. 14, no. Pt 1, pp. 581–588, 2011. [PubMed: 22003665]
- [18]. Somayajula S, Panagiotou C, Rangarajan A, Li Q, Arridge SR, and Leahy RM, “PET image reconstruction using information theoretic anatomical priors,” *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 537–549, 3 2011. [PubMed: 20851790]
- [19]. Ehrhardt MJ, Markiewicz P, Liljeroth M, Barnes A, Kolehmainen V, Duncan JS, Pizarro L, Atkinson D, Hutton BF, Ourselin S, Thielemans K, and Arridge SR, “PET Reconstruction With an Anatomical MRI Prior Using Parallel Level Sets,” *IEEE Trans. Med. Imaging*, vol. 35, no. 9, pp. 2189–2199, 9 2016. [PubMed: 27101601]
- [20]. Wang G and Qi J, “Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization,” *IEEE Trans. Med. Imaging*, vol. 31, no. 12, pp. 2194–2204, 12 2012. [PubMed: 22875244]
- [21]. Kim K, Son YD, Bresler Y, Cho ZH, Ra JB, and Ye JC, “Dynamic PET reconstruction using temporal patch-based low rank penalty for ROI-based brain kinetic analysis,” *Phys. Med. Biol.*, vol. 60, no. 5, pp. 2019–2046, 3 2015. [PubMed: 25675392]
- [22]. Wang G and Qi J, “Edge-preserving PET image reconstruction using trust optimization transfer,” *IEEE Trans. Med. Imaging*, vol. 34, no. 4, pp. 930–939, 4 2015. [PubMed: 25438302]
- [23]. Meltzer CC, Leal JP, Mayberg HS, Wagner HN, and Frost JJ, “Correction of PET data for partial volume effects in human cerebral cortex by MR imaging,” *J. Comput. Assist. Tomogr.*, vol. 14, no. 4, pp. 561–570, Jul-Aug 1990. [PubMed: 2370355]
- [24]. Müller-Gärtner HW, Links JM, Prince JL, Bryan RN, McVeigh E, Leal JP, Davatzikos C, and Frost JJ, “Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects,” *J. Cereb. Blood Flow Metab.*, vol. 12, no. 4, pp. 571–583, 7 1992. [PubMed: 1618936]
- [25]. Soret M, Bacharach SL, and Buvat I, “Partial-volume effect in PET tumor imaging,” *J. Nucl. Med.*, vol. 48, no. 6, pp. 932–945, 6 2007. [PubMed: 17504879]
- [26]. Thomas BA, Erlandsson K, Modat M, Thurfjell L, Vandenberghe R, Ourselin S, and Hutton BF, “The importance of appropriate partial volume correction for PET quantification in Alzheimer’s disease,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 38, no. 6, pp. 1104–1119, 6 2011. [PubMed: 21336694]
- [27]. Bousse A, Pedemonte S, Thomas BA, Erlandsson K, Ourselin S, Arridge S, and Hutton BF, “Markov random field and Gaussian mixture for segmented MRI-based partial volume correction in PET,” *Phys. Med. Biol.*, vol. 57, no. 20, p. 6681, 2012. [PubMed: 23023073]
- [28]. Van Cittert PH, “Zum Einfluß der Spaltbreite auf die Intensitätsverteilung in Spektrallinien II,” *Z. Phys.*, vol. 69, no. 5-6, pp. 298–308, 5 1931.
- [29]. Richardson WH, “Bayesian-based iterative method of image restoration,” *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, 1 1972.
- [30]. Lucy LB, “An iterative technique for the rectification of observed distributions,” *Astron. J.*, vol. 79, p. 745, 6 1974.
- [31]. Yan J, Lim JC, and Townsend DW, “MRI-guided brain PET image filtering and partial volume correction,” *Phys. Med. Biol.*, vol. 60, no. 3, pp. 961–976, 2 2015. [PubMed: 25575248]
- [32]. Song T-A, Yang F, Chowdhury S, Kim K, Johnson K, El Fakhri G, Li Q, and Dutta J, “PET image deblurring and super-resolution with an MR-based joint entropy prior,” *IEEE Trans Comput Imaging*, 2019.
- [33]. Nasrollahi K and Moeslund T, “Super-resolution: A comprehensive survey,” *Mach. Vis. Appl.*, vol. 25, pp. 1423–68, 2014.
- [34]. Wallach D, Lamare F, Kontaxakis G, and Visvikis D, “Super-resolution in respiratory synchronized positron emission tomography,” *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 438–448, 2012. [PubMed: 21997249]



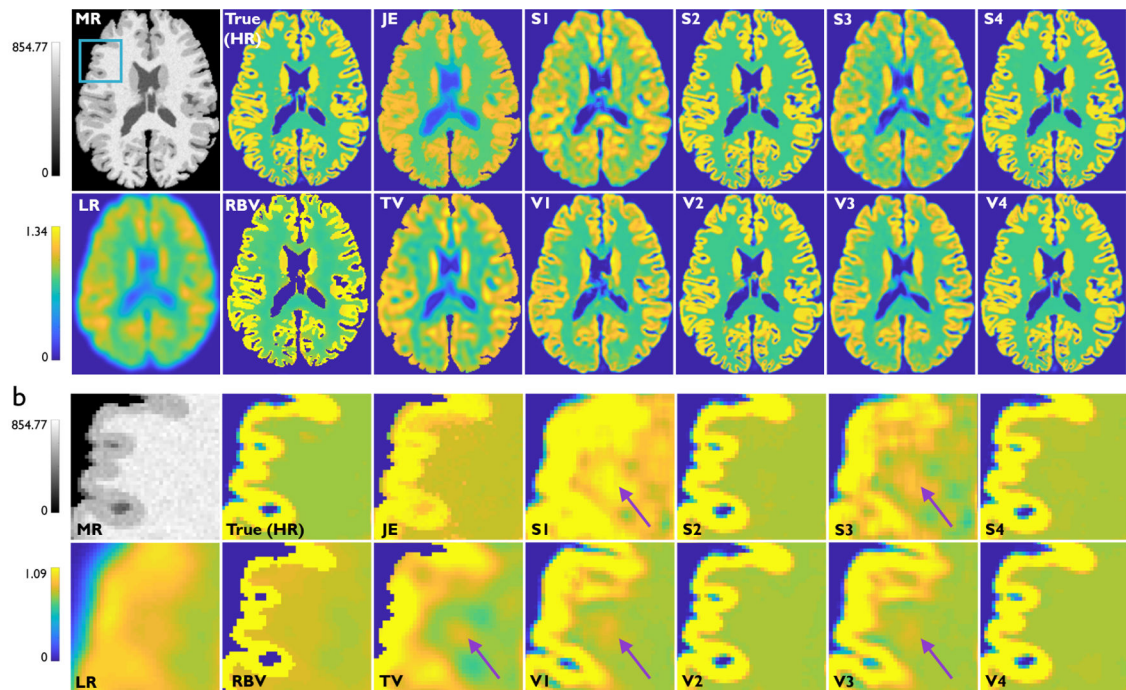
- [35]. Hu Z, Wang Y, Zhang X, Zhang M, Yang Y, Liu X, Zheng H, and Liang D, “Super-resolution of PET image based on dictionary learning and random forests,” *Nucl. Instrum. Methods Phys. Res. A*, vol. 927, pp. 320 – 329, 2019.
- [36]. Cui Z, Chang H, Shan S, Zhong B, and Chen X, “Deep network cascade for image super-resolution,” in *Computer Vision – ECCV 2014*, Fleet D, Pajdla T, Schiele B, and Tuytelaars T, Eds. Cham: Springer International Publishing, 2014, pp. 49–64.
- [37]. Freedman G and Fattal R, “Image and video upscaling from local self-examples,” *ACM Trans. Graph*, vol. 30, no. 2, pp. 12:1–12:11, 4 2011.
- [38]. Glasner MID, Bagon S, “Super-resolution from a single image,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 349–356, 2009.
- [39]. Huang J, Singh A, and Ahuja N, “Single image super-resolution from transformed self-exemplars,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2015, pp. 5197–5206.
- [40]. Chang H, Yeung D-Y, and Xiong Y, “Super-resolution through neighbor embedding,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, 6 2004, pp. I–I.
- [41]. Kim KI and Kwon Y, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, 6 2010. [PubMed: 20431136]
- [42]. Timofte R, Vincent DS, and Van Gool L, “Anchored neighborhood regression for fast example-based super-resolution,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1920–1927, 2013.
- [43]. Yang J, Wright J, Huang T, and Ma Y, “Image super-resolution as sparse representation of raw image patches,” 6 2008.
- [44]. Yang J, Wright J, Huang TS, and Ma Y, “Image super-resolution via sparse representation,” *IEEE Trans. Image Process*, vol. 19, no. 11, pp. 2861–2873, 11 2010. [PubMed: 20483687]
- [45]. Dong C, Loy CC, He K, and Tang X, “Image super-resolution using deep convolutional networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016. [PubMed: 26761735]
- [46]. Kim J, Kwon Lee J, and Mu Lee K, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
- [47]. He K, Zhang X, Ren S, and Sun J, “Deep residual learning for image recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 349–56, 2009.
- [48]. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, and Shi W, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 105–114, 2017.
- [49]. Chen KT, Gong E, de Carvalho Macruz FB, Xu J, Boumis A, Khalighi M, Poston KL, Sha SJ, Greicius MD, Mormino E, Pauly JM, Srinivas S, and Zaharchuk G, “Ultra-Low-Dose 18F-Florbetaben Amyloid PET Imaging Using Deep Learning with Multi-Contrast MRI Inputs,” *Radiology*, vol. 290, no. 3, pp. 649–656, 3 2019. [PubMed: 30526350]
- [50]. Hong X, Zan Y, Weng F, Tao W, Peng Q, and Huang Q, “Enhancing the Image Quality via Transferred Deep Residual Learning of Coarse PET Sinograms,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2322–2332, 10 2018. [PubMed: 29993685]
- [51]. Liu CC and Qi J, “Higher SNR PET image prediction using a deep learning model and MRI image,” *Phys. Med. Biol.*, vol. 64, no. 11, p. 115004, 5 2019. [PubMed: 30844784]
- [52]. da Costa-Luis CO and Reader AJ, “Deep learning for suppression of resolution-recovery artefacts in MLEM PET image reconstruction,” in *IEEE Nucl. Sci. Symp. Conf. Rec.*, 10 2017, pp. 1–3.
- [53]. Song T-A, Chowdhury S, Kim K, Gong K, El Fakhri G, Li Q, and Dutta J, “Super-resolution PET using a very deep convolutional neural network,” in *IEEE Nucl. Sci. Symp. Conf. Rec. IEEE*, 2018.
- [54]. Dutta J, El Fakhri G, Zhu X, and Li Q, “PET point spread function modeling and image deblurring using a PET/MRI joint entropy prior,” in *Proc. IEEE Int. Symp. Biomed. Imaging. IEEE*, 2015, pp. 1423–1426.



- [55]. Cloquet C, Sureau FC, Defrise M, Van Simaeys G, Trotta N, and Goldman S, “Non-Gaussian space-variant resolution modelling for list-mode reconstruction,” *Phys. Med. Biol.*, vol. 55, no. 17, pp. 5045–5066, 9 2010. [PubMed: 20702921]
- [56]. Zhao H, Gallo O, Frosio I, and Kautz J, “Loss functions for neural networks for image processing,” *CoRR*, vol. abs/1511.08861, 2015 [Online]. Available: <http://arxiv.org/abs/1511.08861>
- [57]. Kingma D and Ba J, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [58]. van Velden FH, Kloet RW, van Berckel BN, Buijs FL, Luurtsema G, Lammertsma AA, and Boellaard R, “HRRT versus HR+ human brain PET studies: an interscanner test-retest study,” *J. Nucl. Med.*, vol. 50, no. 5, pp. 693–702, 5 2009. [PubMed: 19372482]
- [59]. Lange K, “Convergence of EM image reconstruction algorithms with Gibbs smoothing,” *IEEE Trans. Med. Imaging*, vol. 9, no. 4, pp. 439–446, 1990. [PubMed: 18222791]
- [60]. Jenkinson M and Smith S, “A global optimisation method for robust affine registration of brain images,” *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, 6 2001. [PubMed: 11516708]
- [61]. Jenkinson M, Bannister P, Brady M, and Smith S, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825–841, 10 2002. [PubMed: 12377157]
- [62]. Wang Z, Bovik AC, Sheikh HR, and Simoncelli EP, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process*, vol. 13, no. 4, pp. 600–612, 4 2004. [PubMed: 15376593]
- [63]. Kim K, Wu D, Gong K, Dutta J, Kim JH, Son YD, Kim HK, El Fakhri G, and Li Q, “Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting,” *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1478–1487, 6 2018. [PubMed: 29870375]

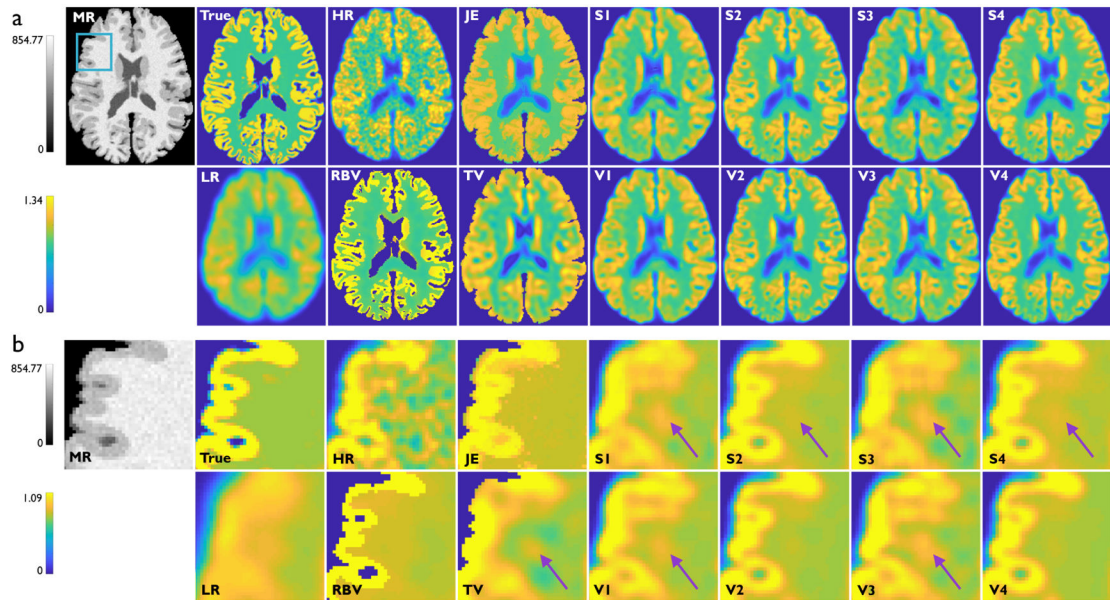


**Fig. 1.** CNN architecture for SR PET. The network uses up to 4 inputs: (i) LR PET (the main input), (ii) HR MR, (iii) radial locations, and (iv) axial locations. The network consists of alternating convolutional (Conv) layers and rectified linear units (ReLUs) for nonlinear activation. It predicts the residual PET image which is later added to the input LR PET to generate the SR PET.



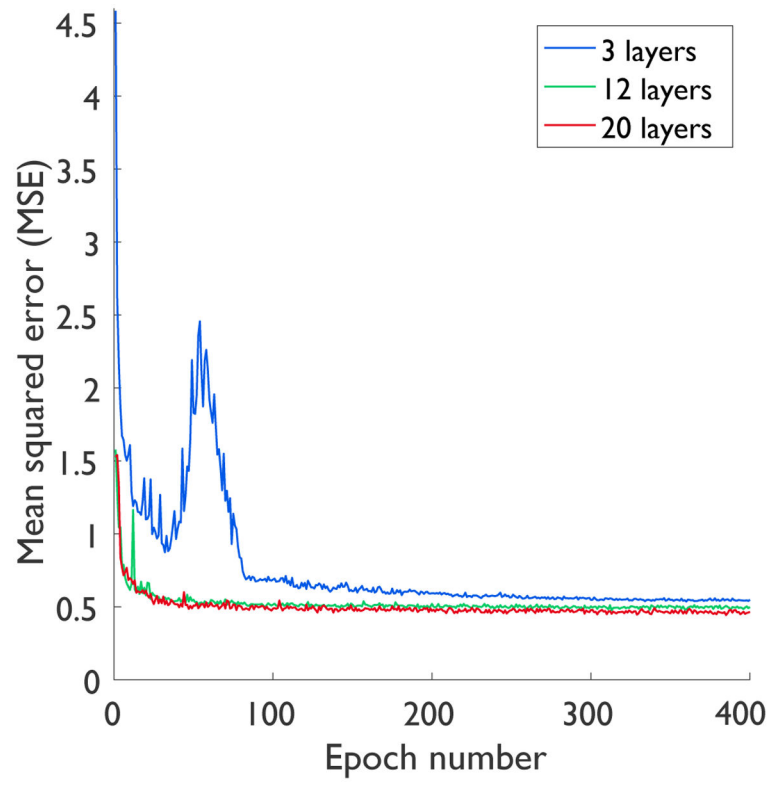
**Fig. 2.**

Study 1: Simulation results from one subject belonging to the validation dataset. (a) Transverse slices from the T1-weighted HR MR image, true PET image (also the HR image for this case), LR PET image (HR+ scanner), RBV-corrected PET image, JE-penalized deconvolution result, TV-penalized deconvolution result, and the  $S_r$  outputs from the following CNNs: S1, V1, S2, V2, S3, V3, S4, and V4. The blue box on the MR image indicates the region that is magnified for closer inspection. (b) The corresponding magnified subimages. Purple arrows indicate areas in the white-matter background region where prominent noise-induced artifacts arise for TV and for the CNNs without anatomical inputs, namely S1, V1, S3, and V3.

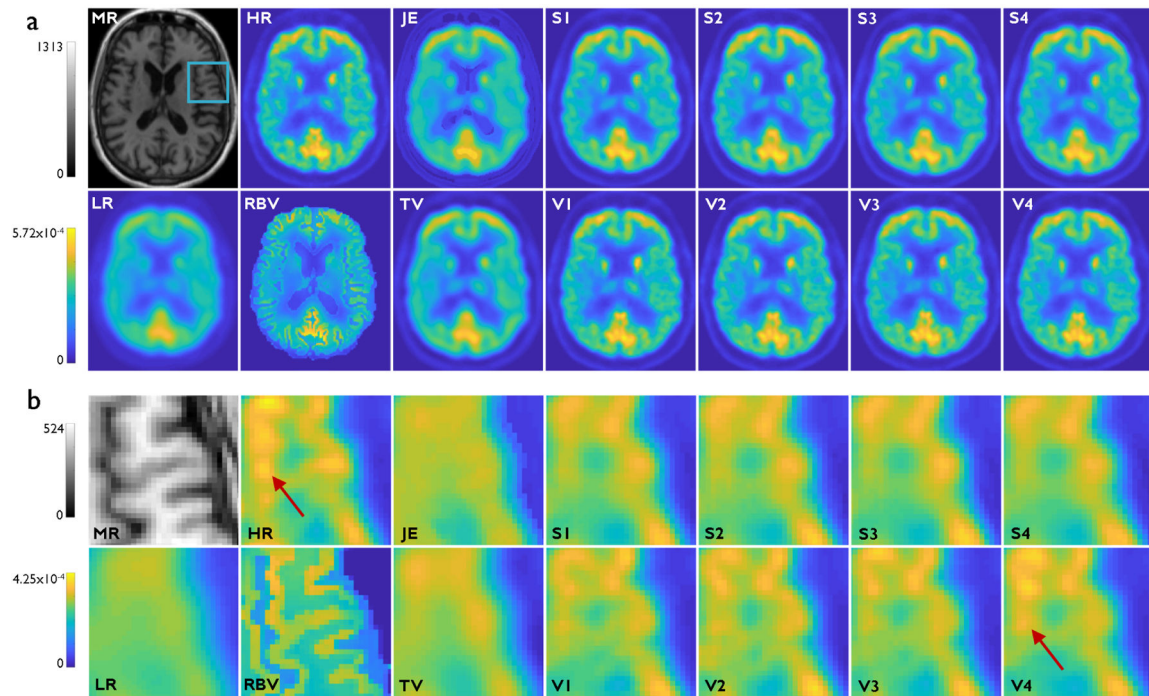


**Fig. 3.**

Study 2: Simulation results from one subject belonging to the validation dataset. (a) Transverse slices from the T1-weighted HR MR image, true PET image, HR PET image (HRRT scanner), LR PET image (HR+ scanner), RBV-corrected PET image, JE-penalized deconvolution result, TV-penalized deconvolution result, and the SR outputs from the following CNNs: S1, V1, S2, V2, S3, V3, S4, and V4. The blue box on the MR image indicates the region that is magnified for closer inspection. (b) The corresponding magnified subimages. Purple arrows indicate areas in the white-matter background region where prominent noise-induced artifacts arise. These artifacts are the least prominent for V2 and V4 — very deep CNNs with anatomical inputs.



**Fig. 4.** Convergence curves for different depths: 3, 12, and 20 layers.



**Fig. 5.**

Study 3: Clinical results from one subject belonging to the validation dataset. (a) Transverse slices from the T1-weighted HR MR image, HR PET image (HRRT scanner), LR PET image (HR+ scanner), RBV-corrected PET image, JE-penalized deconvolution result, TV-penalized deconvolution result, and the Sr outputs from the following CNNs: S1, V1, S2, V2, S3, V3, S4, and V4. The blue box on the MR image indicates the region that is magnified for closer inspection. (b) The corresponding magnified subimages. The very deep CNNs (V1, V2, V3, and V4) yield sharper images than their shallow counterparts (S1, S2, S3, and S4). The red arrows points to bright gray matter areas in the HR image that are recovered with the highest contrast in V4.



**TABLE I**

CNN architectures

Network properties	S1	V1	S2	V2	S3	V3	S4	V4
Number of layers	3	20	3	20	3	20	3	20
Input types	LR PET	LR PET	LR PET HR MR	LR PET HR MR	LR PET Radial locations Axial locations	LR PET Radial locations Axial locations	LR PET HR MR Radial locations Axial locations	LR PET HR MR Radial locations Axial locations



**TABLE II**

Simulation and experimental studies

Study index	Study type	LR image (input)	HR image (target)
1	Simulation	HR+ PET	True PET
2	Simulation	HR+ PET	HRRT PET
3	Clinical	HR+ PET	HRRT PET

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE III**

LR and HR image sources

<b>Image Image</b>	<b>Scanner</b>	<b>Spatial resolution</b>	<b>Bore diameter</b>	<b>Axial length</b>
LR	HR+	4.3 - 8.3 mm	562 mm	155 mm
HR	HRRT	2.3 - 3.4 mm	312 mm	250 mm

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE IV

Study 1: Performance comparison

Metric	Reference	HR	LR	RBV	TV	JE	S1	V1	S2	V2	S3	V3	S4	V4
PSNR	True	$\infty$	21.47	23.68	22.35	22.37	26.86	27.71	37.17	37.61	27.17	27.98	37.27	<b>37.93</b>
PSNR s.d.		0.00	0.41	0.45	0.66	0.68	0.71	0.72	0.77	0.56	0.74	0.76	0.73	0.61
SSIM	True	1.00	0.74	0.74	0.85	0.86	0.82	0.86	0.917	0.97	0.84	0.87	0.97	<b>0.98</b>
SSIM s.d.		0.000	0.007	0.013	0.012	0.006	0.005	0.003	0.002	0.0008	0.005	0.004	0.001	0.001
CNR	-	0.95	0.09	0.59	0.52	0.53	0.50	0.56	0.89	0.85	0.53	0.57	0.88	<b>0.91</b>
CNR s.d.		0.09	0.04	0.06	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.09	0.09

**TABLE V**

Study 2: Performance comparison

Metric	Reference	HR	LR	RBV	TV	JE	S1	V1	S2	V2	S3	V3	S4	V4
PSNR	HRRT	$\infty$	27.83	26.55	26.68	27.48	35.11	35.69	38.37	38.48	35.48	35.92	38.44	<b>38.69</b>
PSNR s.d.		0.00	0.17	0.45	0.55	0.50	0.36	0.33	0.24	0.26	0.33	0.29	0.28	0.27
PSNR	True	22.15	21.47	21.84	22.35	22.37	22.69	23.06	22.46	23.92	23.07	23.48	23.78	<b>24.29</b>
PSNR s.d.		0.52	0.41	0.70	0.66	0.68	0.070	0.70	0.072	0.70	0.71	0.61	0.71	0.70
SSIM	HRRT	1.00	0.76	0.81	0.83	0.83	0.82	0.80	0.88	0.88	0.85	0.82	0.89	<b>0.90</b>
SSIM s.d.		0.000	0.033	0.012	0.006	0.005	0.050	0.001	0.062	0.0028	0.024	0.002	0.014	0.003
SSIM	True	0.84	0.74	0.74	0.85	0.86	0.77	0.75	0.86	0.86	0.79	0.77	0.86	<b>0.87</b>
SSIM s.d.		0.003	0.007	0.013	0.012	0.006	0.022	0.008	0.020	0.006	0.021	0.007	0.021	0.006
CNR	-	0.67	0.09	0.59	0.52	0.53	0.48	0.50	0.58	0.63	0.46	0.49	0.61	<b>0.64</b>
CNR s.d.		0.06	0.04	0.06	0.08	0.08	0.03	0.03	0.12	0.01	0.04	0.04	0.01	0.04

**TABLE VI**

Depth and width comparison

Metric	# of Filters \ # of Layers		3	12	20
	# of Filters	# of Layers			
PSNR	32		37.72	38.48	<b>38.60</b>
SSIM	32		0.882	0.905	<b>0.907</b>
PSNR	64		38.46	38.55	<b>38.69</b>
SSIM	64		0.905	0.907	<b>0.908</b>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE VII

Study 3: Performance comparison

Metric	Reference	HR	LR	RBV	TV	JE	S1	V1	S2	V2	S3	V3	S4	V4
PSNR	HRRT	$\infty$	27.24	28.33	29.65	30.13	33.46	35.83	34.28	36.38	34.50	36.23	34.52	<b>37.02</b>
PSNR s.d.		0.00	0.59	0.38	0.48	0.45	0.58	0.14	0.78	0.19	0.70	0.17	0.83	0.22
SSIM	HRRT	1.00	0.86	0.70	0.93	0.89	0.95	0.97	0.95	0.97	0.95	0.97	0.95	<b>0.98</b>
SSIM s.d.		0.000	0.017	0.02	0.007	0.011	0.008	0.002	0.010	0.002	0.009	0.002	0.009	0.002
CNR	-	0.46	0.19	0.23	0.26	0.23	0.34	0.35	0.39	<b>0.41</b>	0.35	0.39	0.40	<b>0.41</b>
CNR s.d.		0.11	0.15	0.03	0.09	0.09	0.11	0.13	0.13	0.12	0.12	0.12	0.12	0.12