Feasibility of Deep Learning–Based PET/MR Attenuation Correction in the Pelvis Using Only Diagnostic MR Images

Tyler J. Bradshaw[1], Gengyan Zhao[2], Hyungseok Jang[3], Fang Liu[1], and Alan B. McMillan[1]

[1]Departments of Radiology and [2]Medical Physics, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI; and [3]Department of Radiology, University of California, San Diego, San Diego, CA

Corresponding Author:
Tyler J. Bradshaw, PhD
3252 Clinical Science Center, 600 Highland Ave,
Madison, WI 53792;
E-mail: tbradshaw@wisc.edu

ABSTRACT

This study evaluated the feasibility of using only diagnostically relevant magnetic resonance (MR) images together with deep learning for positron emission tomography (PET)/MR attenuation correction (deepMRAC) in the pelvis. Such an approach could eliminate dedicated MRAC sequences that have limited diagnostic utility but can substantially lengthen acquisition times for multibed position scans. We used axial T2 and T1 LAVA Flex magnetic resonance imaging images that were acquired for diagnostic purposes as inputs to a 3D deep convolutional neural network. The network was trained to produce a discretized (air, water, fat, and bone) substitute computed tomography (CT) (CT_{sub}). Discretized ($CT_{ref-discrete}$) and continuously valued (CT_{ref}) reference CT images were created to serve as ground truth for network training and attenuation correction, respectively. Training was performed with data from 12 subjects. CT_{sub}, CT_{ref}, and the system MRAC were used for PET/MR attenuation correction, and quantitative PET values of the resulting images were compared in 6 test subjects. Overall, the network produced CT_{sub} with Dice coefficients of 0.79 ± 0.03 for cortical bone, 0.98 ± 0.01 for soft tissue (fat: 0.94 ± 0.0; water: 0.88 ± 0.02), and 0.49 ± 0.17 for bowel gas when compared with $CT_{ref-discrete}$. The root mean square error of the whole PET image was 4.9% by using deepMRAC and 11.6% by using the system MRAC. In evaluating 16 soft tissue lesions, the distribution of errors for maximum standardized uptake value was significantly narrower using deepMRAC ($-1.0\% \pm 1.3\%$) than using system MRAC method ($0.0\% \pm 6.4\%$) according to the Brown–Forsy the test ($P < .05$). These results indicate that improved PET/MR attenuation correction can be achieved in the pelvis using only diagnostically relevant MR images.

INTRODUCTION

Accurate correction for photon attenuation remains a challenge for quantitative positron emission tomography (PET)/magnetic resonance (MR) imaging. Owing to the absence of transmission imaging during PET/MR imaging, either through computed tomography (CT) or a transmission source, MR-based attenuation correction (MRAC) methods are needed to estimate the pixel-wise photon attenuation coefficients for quantitative PET reconstruction. Many MRAC techniques have been developed over the past decade (1, 2), with dual-echo chemical shift-encoded (2-point Dixon) imaging being used in most commercial PET/MR scanners. In Dixon-based MRAC, a single acquisition yields images that are separated into water and fat components and then assigned Hounsfield units (HU) for air, fat, lung, and water (3, 4). With Dixon-based methods, however, bone is not identi-

fied owing to the lack of signal contrast between bone and air (5). While ignoring bone in MRAC appears to have little impact on the diagnostic accuracy of PET/MR imaging (6), it can lead to quantitative PET errors exceeding 20%, depending on the location (7).

Numerous approaches have recently been proposed for localizing bone in MR images, including ultrashort echo time (UTE)– based methods, zero echo time (ZTE)– based methods, atlas-based methods, and PET-only methods (8). In general, these advanced methods have allowed for bone localization with acceptable accuracy, reducing quantitative PET errors to within 5%. However, most advanced MRAC methods have been developed specifically for brain imaging. MRAC outside of the brain, especially in the pelvis, is made much more difficult by the greater variety of tissue types, shapes, and tissue deformation

present in the images. The few methods accounting for bone that have been tested outside of the brain include atlas methods (9–11), a UTE-based method tested in the neck (12), and PET-only methods tested in the upper body (13, 14). Despite encouraging results, each method has drawbacks, such as the need for additional dedicated MRAC scanning or reliance on image registration, and only a few methods have been tested in the pelvis. Further developments in MRAC methods for the pelvis are therefore needed.

Deep learning methods, particularly convolutional neural networks (CNNs), have recently achieved remarkable success in performing complex computer vision tasks and are now being adapted into a broad range of medical imaging applications (15, 16). We previously used a convolutional encoder-decoder (CED) network for PET/MR attenuation correction in the brain, where contrast-enhanced T1-weighted MR images were used as network inputs to achieve reconstructed PET errors of ~1% in the brain (17). We also evaluated the same CED network but with a UTE image as input and using transfer learning to initialize the network weights (18). This method achieved even better results, with PET errors in the brain generally <1%. Recently, Leynes et al. used deep learning to synthesize CT images for PET/MR imaging in the pelvis. The model was trained with both ZTE and Dixon-based images from 10 subjects as inputs with coregistered CT images as ground truth, and significantly improved accuracy was achieved (19). A drawback of this method is the additional scanning time needed for ZTE imaging, which can substantially lengthen the overall scan time for a multibed position acquisition yet has little diagnostic utility. In addition, bowel gas was ignored (filled in with soft tissue HU) owing to the challenge in coregistering reference CT images to MR images. We hypothesize that clinically relevant MR images can be used as input to a deep learning model, eliminating the need for dedicated MRAC sequences such as UTE and ZTE. These sequences can then be used for both diagnostic and attenuation-correction purposes, improving clinical workflow for whole-body imaging, and may be easier to harmonize across scanners and body regions than complex UTE/ZTE imaging.

In this work, we assessed a deep learning–based attenuation-correction method (deepMRAC) in PET/MR imaging of the pelvis that uses diagnostically useful MR images as input and therefore does not require dedicated MRAC imaging. Also, owing to deficiencies that we observed in using CT-to-MR registration to generate reference attenuation maps, we developed a novel method for creating reference attenuation maps for training our network that allows for the segmentation of bowel gas. We evaluated our method's reconstructed PET error relative to the reference attenuation map and the scanner's standard MRAC method in subjects who underwent pelvic PET/MR examinations.

METHODOLOGY

The retrospective study was approved by the institutional review board, and the need for written informed consent was waived. All research was conducted in accordance with the Health Insurance Portability and Accountability Act.

Subjects and Imaging

Subjects were eligible for inclusion in our study if they underwent a clinical ^{18}F-fluorodeoxyglucose (FDG) PET/CT examination and an FDG PET/MR examination at our institution on a 3T Signa PET/MR scanner (GE Healthcare, Waukesha, WI) for evaluation of disease in the pelvis. The PET/MR scan must have been conducted immediately following the PET/CT examination so that a reference CT image was available for analysis. Subjects must also have had T2 MRI, T1 LAVA Flex, and system MRAC images acquired during the PET/MR scan with sufficient transaxial field of view (FOV) to cover the entire pelvis. In total, 18 female subjects (mean age, 63 years; range, 39–89 years), all with cervical cancer, met these criteria and were included in the study. Subjects were injected with 0.14 mCi/kg of FDG according to our institution's PET imaging protocol. Because PET/MR imaging was conducted after a PET/CT scan, PET imaging was conducted an average of 135 ± 25 minutes after radiotracer injection.

For each bed position of the whole-body PET/MR scan, the following images were acquired: PET (3 min/bed position), the system MRAC acquisition using the body transmit–receive coil as well as axial T2 fast recovery fast spin echo, and T1 LAVA Flex using whole-body array coils. In all subjects, gadolinium contrast was administered (gadobenate dimeglumine [Multi-Hance; Bracco Diagnostics, Princeton, NJ] at a dosage of 0.1 mmol/kg) during a dedicated single bed-position acquisition 17 ± 4 minutes before whole-body imaging, in which case, the contrast was mostly diluted by the time whole-body PET/MR imaging was performed. Each image was inspected and no fat/water swapping was observed in the MRAC acquisition. The following parameters for the T2 acquisition were included: FOV = 500 mm, pixel size = 0.94×0.94 mm^2, echo time = 87–100 milliseconds, repetition time = 4500 milliseconds, and section thickness = 6 mm. The following parameters for the T1 LAVA Flex were included: FOV = 500 mm, pixel size = 0.94×0.94 mm^2, TE1/TE2 = 1.35/2.04 milliseconds, repetition time = 5.4 milliseconds, and section thickness = 2 mm.

CT images used in this study were acquired as part of the PET/CT scan occurring immediately before the PET/MR examination. PET/CT scans were acquired on either a GE Discovery 710 or a Discovery VCT PET/CT scanner. CTs were acquired with the following parameters: voltage = 140 kVp, automatic exposure control with a noise index = 25, rotation time = 0.5 second, pitch = 0.516, and section thickness = 5 mm with intersection spacing = 3.27-mm. Note that these CT images were of higher image quality than typical low-dose attenuation-correction CT images, as these are read separately for diagnostic purposes at our institution. Note that it is unclear if typical lose-dose attenuation-correction CT images could have instead been used in the study with equivalent results.

Reference CT Generation

The validation of MRAC methods requires the availability of ground truth attenuation maps representing the anatomy *at the time of PET/MR imaging*. Nearly all previous studies evaluating MRAC methods have used coregistered CT images as ground truth (9, 19). The pelvis images used in our study posed a challenge for existing multimodal registration algorithms, as
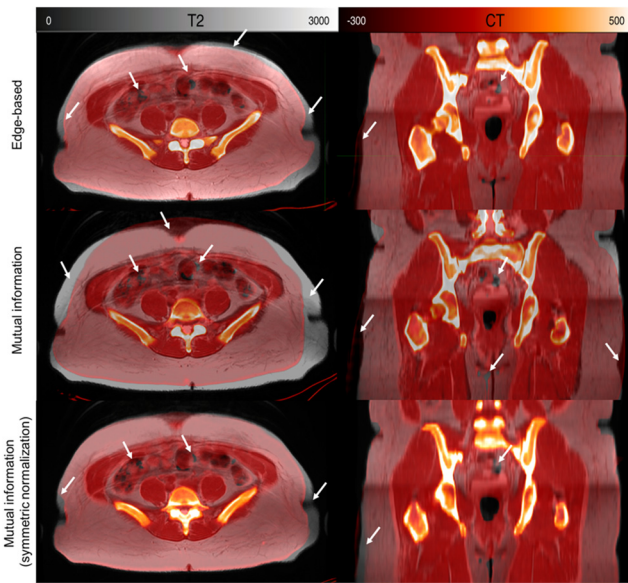
Figure 1. Postregistration fused magnetic resonance (MR)/computed tomography (CT) images show the limitations of using a coregistered CT as the reference for evaluating MRAC methods in the pelvis. Edge-based deformable registration (*top*), mutual information–based deformable registration (*middle*), and mutual information deformable registration with symmetric normalization (*bottom*) result in residual mismatch for soft tissue and especially bowel gas, as indicated by the white arrows.

there were differences in body positioning (eg, from straight to bent legs, curved to flat couch), bowel gas location (eg, pockets of bowel gas), arm location, and organ location/shape (eg, bladder filling or bowel movement) between the times of MR acquisition and CT. We tested multiple commercial and open-source deformable image registration packages on our pelvis data, including algorithms that were used in previous studies. We found that although some algorithms performed better than others, all had minor and sometimes large soft tissue and bone misregistrations, and none performed well at registering bowel gas (Figure 1). Previous studies have ignored bowel gas by filling in air bubbles with tissue-equivalent CT numbers (19), although the effect of this technique on PET reconstruction error is unknown. We felt these misregistrations could compromise the deep learning network's ability to learn pixel-to-pixel MR-to-CT mapping. For these reasons, we chose to not use image registration as the sole means of obtaining ground truth attenuation maps, but to instead synthesize a reference CT by using a combination of different techniques for different tissue types (fat/water, bone, and air), as illustrated in Figure 2.

Fat and water localization was derived from the subject's MRI images acquired in the same session. The subject's fat-only and water-only images from the LAVA Flex acquisition were converted into a fat fraction image according to the following equation:

$$FF_i = \frac{F_i}{(F_i + W_i)},$$

where FF_i is the fat fraction value for voxel i (ranging from 0 to 1), F_i is the signal intensity of voxel i in the fat-only image, and W_i is the signal intensity of voxel i in the water-only image (20). The fat fraction image was then converted into Hounsfield unit (HU) using the following relationship:

$$HU_i = (1 - FF_i)HU_{water} + FF_iHU_{fat},$$

where HU_{water} is the HU of water and HU_{fat} is the HU of fat. We determined that using 55 HU for HU_{water} and −115 HU for HU_{fat} resulted in fat and water having the same average HU as in the acquired CT image. This was determined by performing linear regression between HU values measured in the CT and in the fat fraction images in several subjects. Note that because FF_i in water and fat approach but never actually reach 0 or 1, the assigned HU_i never actually reached 55 for water or −115 for fat. Instead, the resulting images had mean HU values of approximately −100 in fat and 42 in water.

Because of the rigidity of bone, we used image registration for bone localization. A subject's CT image was registered to the T2 MR image, first using manual registration followed by deformable registration (to account for differences in rotation/bending and MR spatial distortions). Registrations were performed using the CT-MR registration algorithm in Mirada XD (Mirada, Oxford, UK), which is based on mutual information with radial basis functions. In several cases, automatic registration resulted in inadequate bone registration. In these cases, the degree of deformation in the registration was reduced/smoothed (an option in the software) or the images were manually aligned in the software on the basis of visual assessment. For a few cases in which the previously described steps were deemed insufficient, bone-by-bone registration was performed, where bones were individually segmented and independently registered in Mirada. Following registration, the bones in the pelvic region (pelvis, spine, sacrum, coccyx, and femurs) were then extracted from the registered CT images using an in-house atlas-based CT segmentation algorithm (21). The bones were then transplanted to the reference CT image, resulting in a combined fat, water, and bone synthetic CT image.

Air, including bowel gas, was localized by using an intensity-based threshold on the axial T2 image followed by morphological closing and manual adjustment. The threshold was defined as 2 standard deviations below the mean intensity value of at least 3 regions of interest placed in different muscles. Morphological closing was performed in MATLAB 2016a (MathWorks, Natick, MA) using a 4-voxel-wide structuring element to remove noisy voxels that fell below the threshold. We corrected for differences in MRI intensity occurring between different bed positions by normalizing the images at each bed position by the mean intensity value in the muscle. Additional low-signal tissues that were below the intensity threshold (eg, large blood vessels, certain muscles) and consequently labeled as air were visually identified and manually labeled as tissue-equivalent. Bowel gas regions of interest were then translated to the reference CT image and assigned an HU of −1000. The resulting continuously valued substitute CT image, representing the anat-
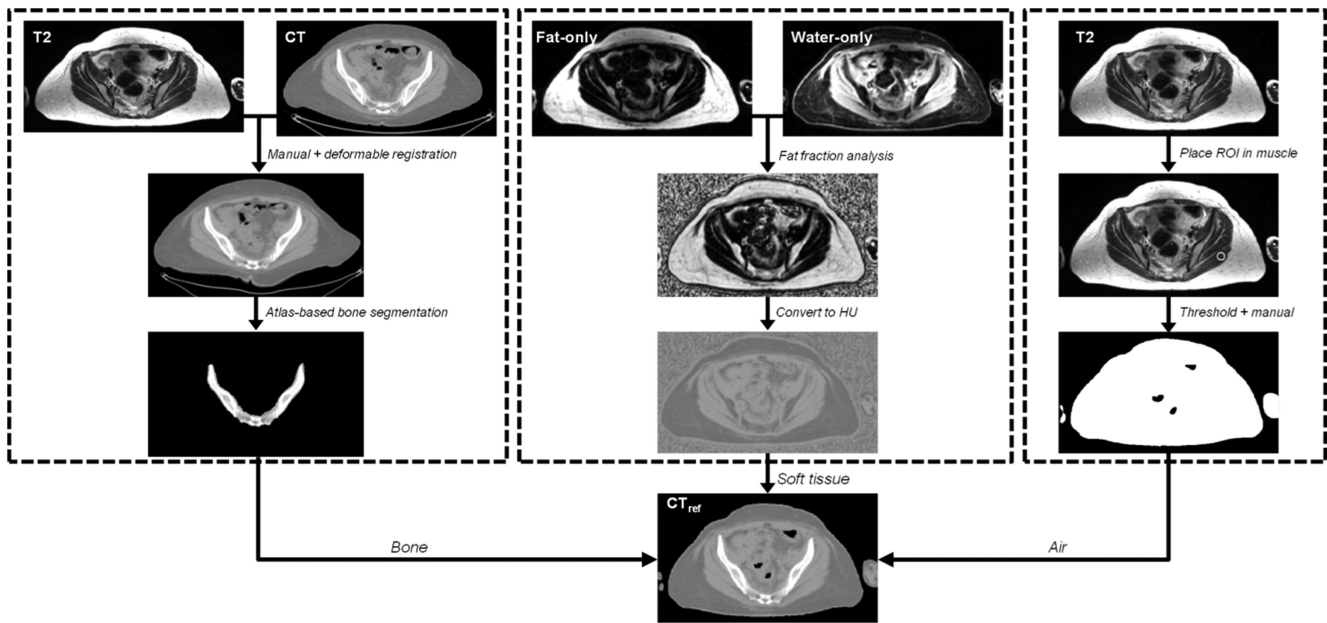
Figure 2. Reference CT (CT$_{ref}$) images were generated to train the network and evaluate MR attenuation correction (MRAC) accuracy. Bone for CT$_{ref}$ was generated via registration of the subject's CT image with their T2 MR image followed by segmentation of the bone. Fat and water components were determined from a fat-fraction image, generated from the 2-point Dixon acquisition. Air, including bowel gas, was identified using an intensity threshold of the T2 image on the basis of a region of interest (ROI) in the muscle, with manual corrections.

omy at the time of MR acquisition, served as the reference CT (CT$_{ref}$) and was used as the ground truth to evaluate PET reconstruction error.

Network Architecture

In this study, we used the open-source multiscale 3D CNN developed by Kamnitsas et al. (22) (Figure 3) as the core of our deepMRAC method. The network, known as DeepMedic, was originally developed to segment lesions in brain MR images. It

achieved high rankings in the BraTS 2015 brain tumor segmentation challenge (23) and in the ISLES 2016 ischemic stroke lesion segmentation challenge (24). The network processes an image using an efficient patch-based method on the basis of image segments, but uses two 3D segments per voxel—a large-scale segment for contextual awareness and a small-scale segment for fine detail. The 2 segments are simultaneously processed by 2 independent CNNs, which get combined at the end via fully connected layers. The model has
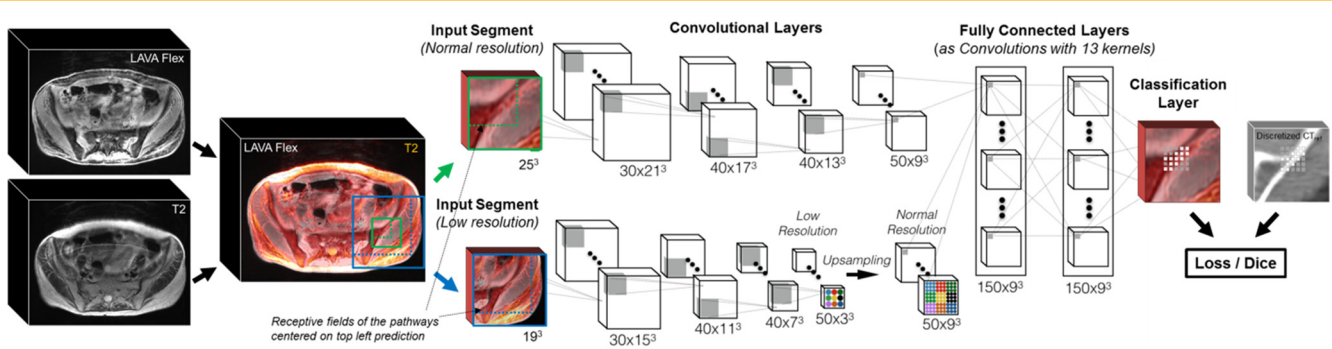

Figure 3. The DeepMedic network architecture is shown, adapted for use with pelvis MR images (T2 and LAVA Flex water-only images). The network uses two 3D convolutional neural network (CNN) pathways, each with a different receptive field relative to the input image, trained in parallel, and then combined via fully connected layers. The model's output is a 4-class probability map, which is then compared with the discretized reference CT. Not all of DeepMedic's 11 layers are shown in the figure for illustration purposes. This figure was adapted, with permission, from Kamnitsas et al. (22).

been made available at https://biomedia.doc.ic.ac.uk/soft-ware/deepmedic/.

Model Training and Testing

The 18 subjects were randomly split into 12 training subjects and 6 testing subjects. For each subject, the input to the model was the paired T2 and T1 LAVA Flex water-only images stacked as image channels, both of which are acquired for diagnostic use at our institution. The model was trained to produce a discretized version of the CT_{ref} ($CT_{ref-discrete}$), where the continuously valued CT_{ref} image was converted into a 4-class mask of air, fat, water, and cortical bone for training. The continuously valued CT_{ref} was discretized using the following thresholds: air < -200 HU, -200 HU $\leq$ fat < -20 HU, -20 HU $\leq$ water < 125 HU, cortical bone ≥ 125 HU. These thresholds were selected because they resulted in discretized classes with visual patterns that were consistent across subjects, which made them easier for the model to learn. Once trained, the model uses the LAVA Flex and T2 images to produce 4-class probability maps with the same dimensions as the input images. Tissue masks are created by assigning each voxel to the tissue with the highest class probability. All images were resampled to 256×256 (voxel dimensions = $1.95 \times 1.95 \times 3$ mm³) before training/testing, with the axial FOV cropped to include only the pelvis. Training was performed on batches of 1500 3D image segments of $25 \times 25 \times 25$ using the Theano library. Optimization was performed using RMSprop optimization with a learning rate of 0.001 and a momentum of 0.6 with a multiclass cross entropy loss function. Training was run for 35 epochs on an NVIDIA 1080 Ti graphics processing unit (GPU).

We experimented with various data augmentation techniques, such as enlarging the image by 10% (ie, making the voxel sizes smaller), flipping the images left to right, and rotating the images in various directions by 5°. Overall, results were improved by only flipping the images left to right. Rotation and image zoom were therefore not used during training. We also tested if an ensemble of 3 DeepMedic models trained using the same data and same network but with randomized training order and initial model weights could improve test results. Ensemble methods have gained in popularity and recently placed first in the 2017 BraTS brain segmentation challenge (25). The probability maps for each tissue class were summed across the 3 models, and the voxel was assigned to the class with the highest summed probability. In addition, we tested if training with a single input series (ie, T2-only or LAVA Flex-only) was as successful as training with both input series.

Testing was performed on images from 6 subjects not used in the training phase. The model's output for each of the 6 subjects was compared with the $CT_{ref-discrete}$ images using Dice similarity coefficients (DSC) for each tissue class.

PET Reconstruction and Analysis

For the 6 testing subjects, we converted the model's output (4-class mask) into a substitute CT image (CT_{sub}) so it could be used for attenuation correction in PET/MR image reconstruction. This was accomplished by assigning an HU of -1000 to air, -100 to fat, 42 to water, and 300 to cortical bone. Fat and water HU were determined by measuring their mean HU values in acquired CT images. Because of the large variation in HU of cortical bone and its nonlinear impact on attenuation, bone HU was empirically determined by testing several HU values (200, 300, 400, and 500) during reconstruction and finding the value that minimized PET reconstruction error.

PET images were reconstructed using both the continuously valued CT_{ref} (ground truth) and CT_{sub} for attenuation and scatter correction. In addition, PET images were reconstructed using the system's default MRAC method on the basis of 2-point Dixon—a method that does not account for bone. Reconstructions were performed using an offline reconstruction toolbox (PET Toolbox, GE Healthcare) with the following PET reconstruction settings: iterations = 4, subsets = 28, transaxial postfilter = 4 mm, FOV = 600 mm, and voxel dimensions = $2.34 \times 2.34 \times 2.78$ mm³. Two PET bed positions were reconstructed, which sufficiently covered the pelvis because of the Signa PET/MR's large axial FOV.

Sixteen FDG-avid soft tissue lesions were identified in the PET images of the 5 subjects on the basis of radiology reports (1 subject had no PET-avid disease), consisting of 7 retroperitoneal lymph nodes, 5 lesions along the vaginal cuff, 1 cervical lesion, 1 ovarian lesion, 1 periaortic lymph node, and 1 pelvis sidewall lymph node. Lesions were contoured using PET Edge in MIM (MIM Software Inc, Cleveland, OH), from which maximum standardized uptake values (SUV_{max}) and mean SUV (SUV_{mean}) were measured. Errors in PET SUVs were determined for the deepMRAC method and those for the system MRAC by comparing SUVs against those reconstructed with CT_{ref}. Because PET SUV errors were both positively and negatively centered around 0, the variances of the distributions of SUV errors were compared using the Brown–Forsythe test, with a $P < .05$ significance level. We also computed the root mean square error (RMSE) of the entire PET image (excluding low-signal voxels $<$ 500 Bq/mL) according to the methods described by Ouyang et al. (26), in which:

$$RMSE = \sqrt{\mu^2 + \sigma^2},$$

where μ is the mean voxel bias, and σ is the standard deviation of the voxel bias (bias being defined as the percent difference relative to the reference PET image).

Sensitivity Analysis

Because deepMRAC produces a discretized CT instead of a continuously valued CT for attenuation correction, we evaluated the effect of using a discrete CT instead of a continuously valued CT for PET attenuation correction. We reconstructed the 6 test subjects' PET images using $CT_{ref-discrete}$ and calculated the PET error ($RMSE_{disrete}$) and lesion SUV error relative to using CT_{ref}.

We also evaluated the impact of ignoring bowel gas on attenuation correction. We filled in bowel gas with tissue equivalence (HU = 42) in the CT_{ref}, and we calculated the PET error ($RMSE_{gas}$) and lesion SUV error relative to using the original CT_{ref} (note that RMSE is calculated only for voxels of >500 Bq/mL and therefore ignored the error *within* the gas pockets).

PET/MR attenuation correction with CT_{ref} was also compared against a reference CT generated using deformable registration alone, with the methods and results reported in the online Supplemental Material.

Table 1. Comparison of DeepMedic's Tissue Segmentations Against the Discretized Reference CT for 6 Testing Subjects

Model Inputs	Average Dice Similarity Coefficient (±SD)					
	Air	Bowel Gas	Fat	Water	Soft Tissue (Fat + Water)	Cortical Bone
T2-only	0.99 ± 0.01	0.43 ± 0.24	0.93 ± 0.02	0.85 ± 0.02	0.98 ± 0.01	0.73 ± 0.05
T1 LAVA Flex-only	0.98 ± 0.01	0.16 ± 0.14	0.92 ± 0.02	0.86 ± 0.03	0.97 ± 0.01	0.73 ± 0.05
Combined	0.99 ± 0.01	0.49 ± 0.17	0.94 ± 0.01	0.88 ± 0.02	0.98 ± 0.01	0.79 ± 0.03
Combined (ensemble)	0.99 ± 0.00	0.48 ± 0.21	0.94 ± 0.01	0.88 ± 0.03	0.98 ± 0.01	0.79 ± 0.03

RESULTS

Training of the network took approximately 8 hours on a single graphics processing unit, while inference on a single subject took 30 seconds. Table 1 shows the DSC comparing the tissue masks segmented via DeepMedic and that of $CT_{ref\text{-}discrete}$ in the testing subjects. When using both T2 and LAVA Flex images as model input, the DSC for fat (0.94 ± 0.01) and water (0.88 ± 0.02) was high. The DSC for cortical bone was slightly lower at 0.79 ± 0.03. The DSC for bowel gas was substantially lower at 0.49 ± 0.17. Using both MRI series as inputs to the model resulted in better DSC for soft tissue, gas, and bone than using only 1 MRI series as input. The ensemble of 3 DeepMedic networks produced, on average, almost identical DSCs as for a single network and, subjectively, did not reduce the presence of small islands of misplaced bone or gas. Figure 4 shows the generated CT_{sub} in comparison to $CT_{ref\text{-}discrete}$, the input T2 series, and the system MRAC attenuation-correction map for an example subject.

Although a large majority of the segmented tissue masks closely resembled the ground truth tissue masks, there were a few apparent errors committed by DeepMedic. The most common error
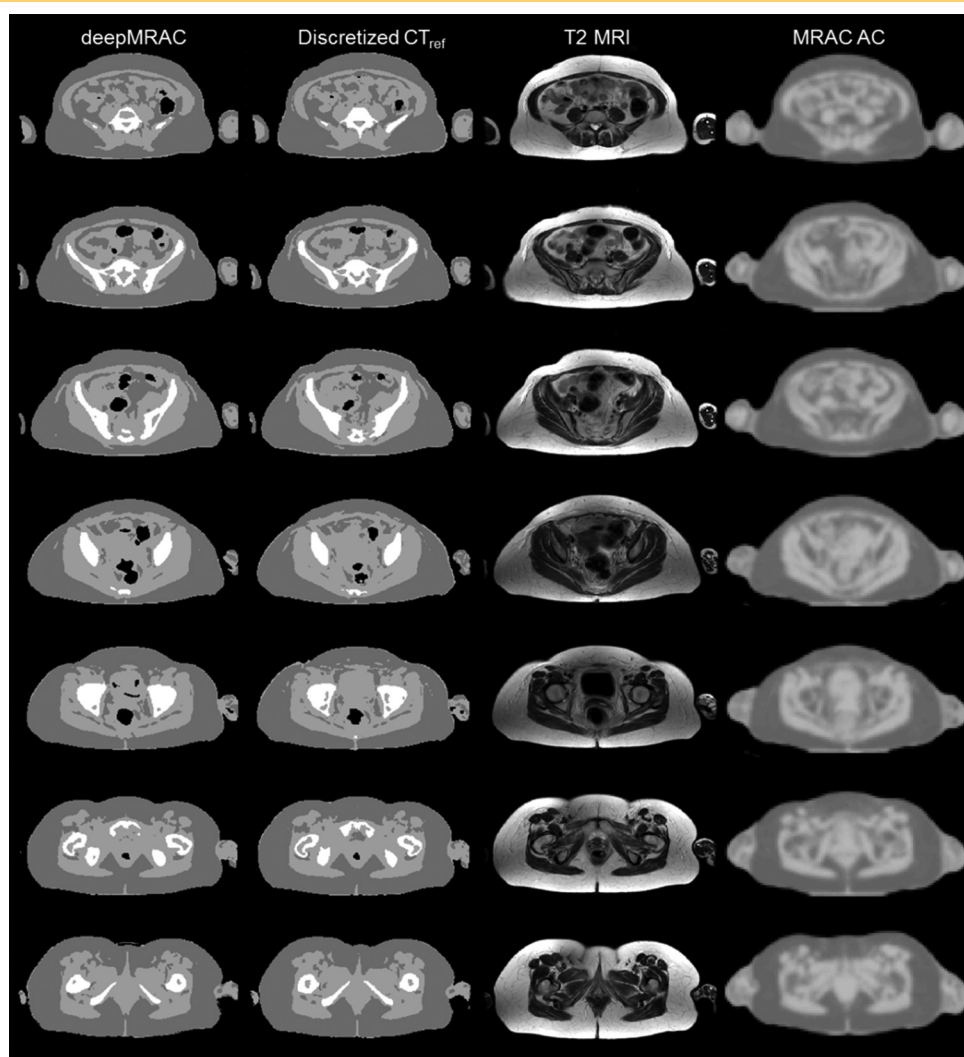

Figure 4. Axial sections from an example subject's model output (*left*), discretized reference CT ($CT_{ref\text{-}discrete}$) (*left-middle*), T2 MR images (*right-middle*), and system MRAC attenuation correction map (*right*). Axial slices were selected at intervals of 3 cm.
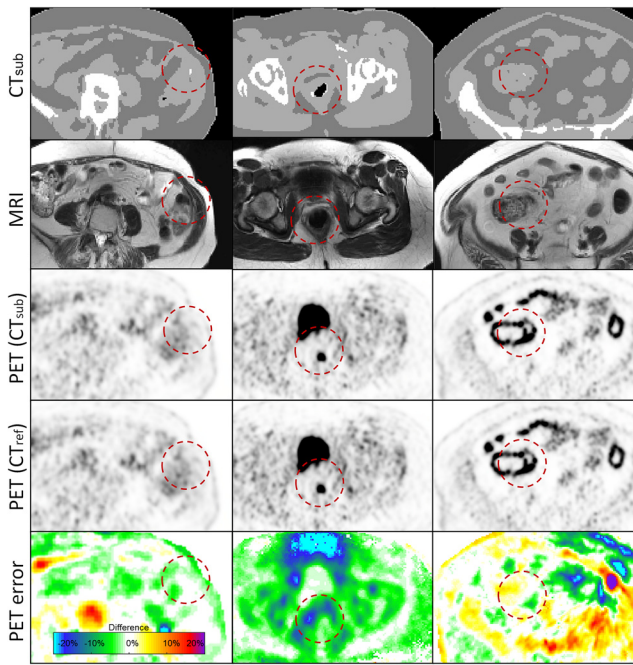
Figure 5. An artifact introduced by DeepMedic is misplaced bone, which occurred in 3 of the 6 testing subjects, all of which are shown here. Also shown are the corresponding T2 magnetic resonance imaging (MRI) images, the positron emission tomography (PET) images reconstructed with CT_{sub}, and the reference PET images reconstructed with CT_{ref}. Note that the small volumes of misplaced bone have negligible impact on the reconstructed PET.

Figure 6. Distributions of standardized uptake value (SUV) errors in 16 soft tissue lesions when using deepMRAC and the system's MRAC method.

was overlooking very thin bones, which were often only 1–2 voxels in width after resampling. This sometimes resulted in discontinuous bone masks. Other errors were more noticeable, including misplacement of bone (occurred in 3 cases, as shown in Figure 5), mistaking parts of the bladder for air (occurred in 1 case), and mistaking MRI ghosting artifacts for tissue (occurred in 1 case). Most of these mistakes were minor and resulted in only 3–15 voxels being incorrectly classified, and had a negligible impact on the resulting PET images (Figure 5).

Figure 6 shows the distribution of SUV errors introduced by using either deepMRAC or the system MRAC in the 16 lesions. The variance of SUV errors in the lesions was significantly smaller when using deepMRAC than when using the system MRAC, for both SUV_{max} ($P = .003$) and SUV_{mean} ($P = .01$). PET images and error maps are shown for an example subject in Figure 7.

The whole-image PET RMSE was found to be 4.9% using deep-MRAC and 11.6% when using the system MRAC. This is illustrated for an example subject in Figure 8, where substitute CT images are shown next to their corresponding voxel-wise PET error maps.

The PET error ($RMSE_{discrete}$) introduced by using a discrete $CT_{ref\text{-}discrete}$ instead of continuously valued CTAC (CT_{ref}) was 4.4%. Lesion SUV errors were $-3.0 \pm 1.3\%$. When filling bowel gas with tissue equivalence, the resulting PET $RMSE_{gas}$ was 5.9%, and lesion SUV errors were $0.4 \pm 3.5\%$.
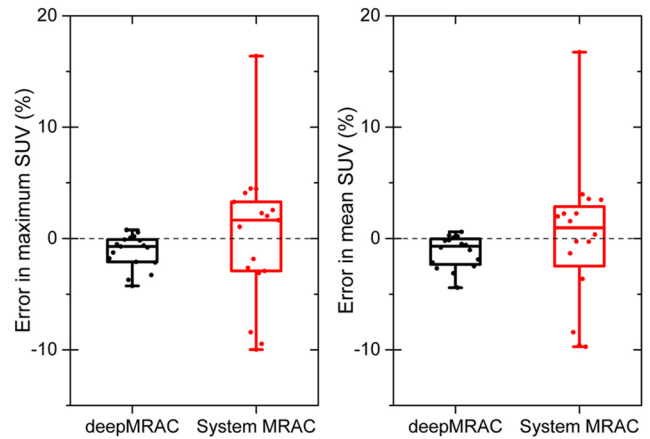
DISCUSSION

PET/MR imaging is a relatively new yet quickly evolving field. The recent surge in PET/MR research has yielded a number of MR-based attenuation-correction methods that have produced impressive results, particularly for brain imaging (8). However, deep learning approaches have certain advantages over other advanced methods. A common criticism of atlas-based methods is the underlying assumption that a subject has a similar overall anatomy as a population model (or similar to one of the atlases for multiatlas approaches). Deep learning only requires that the local image features in a subject resemble features that were observed during training, thus not requiring global similarity in anatomy. This characteristic of deep learning methods also can result in unexpected artifacts (Figure 5), which can likely be mitigated with more and better training data.

Few MRAC methods have been evaluated in PET/MR imaging of the pelvis. Leynes et al. reported on a method based on hybrid ZTE/Dixon imaging, which they evaluated in pelvic images of 6 patients undergoing PET/MR imaging (27). They reported a PET RMSE of 4.18%, which is comparable to deepMRAC's RMSE of 4.9%. In their follow-up study using deep learning with ZTE/Dixon imaging, they reported a PET RMSE of 2.85% (19). A task closely related to MRAC is the use of MR images for radiotherapy treatment planning (MRRT), both requiring the derivation of photon attenuation maps from MR images. Several MRRT planning methods for the pelvis have been reported, including atlas-based methods (28), voxel regression methods (29, 30), and traditional machine learning methods (31). Many of these, however, are not fully automated and require user input (eg, manual bone segmentation) (29, 30). Furthermore, comparison across methods is challenging owing to the different evaluation techniques and definitions. For example, the tissue class of bone consisted of only cortical bone in our method (because bone marrow has attenuation coefficients similar to water), whereas MRRT studies often consider the total bone volume during evaluation (28).
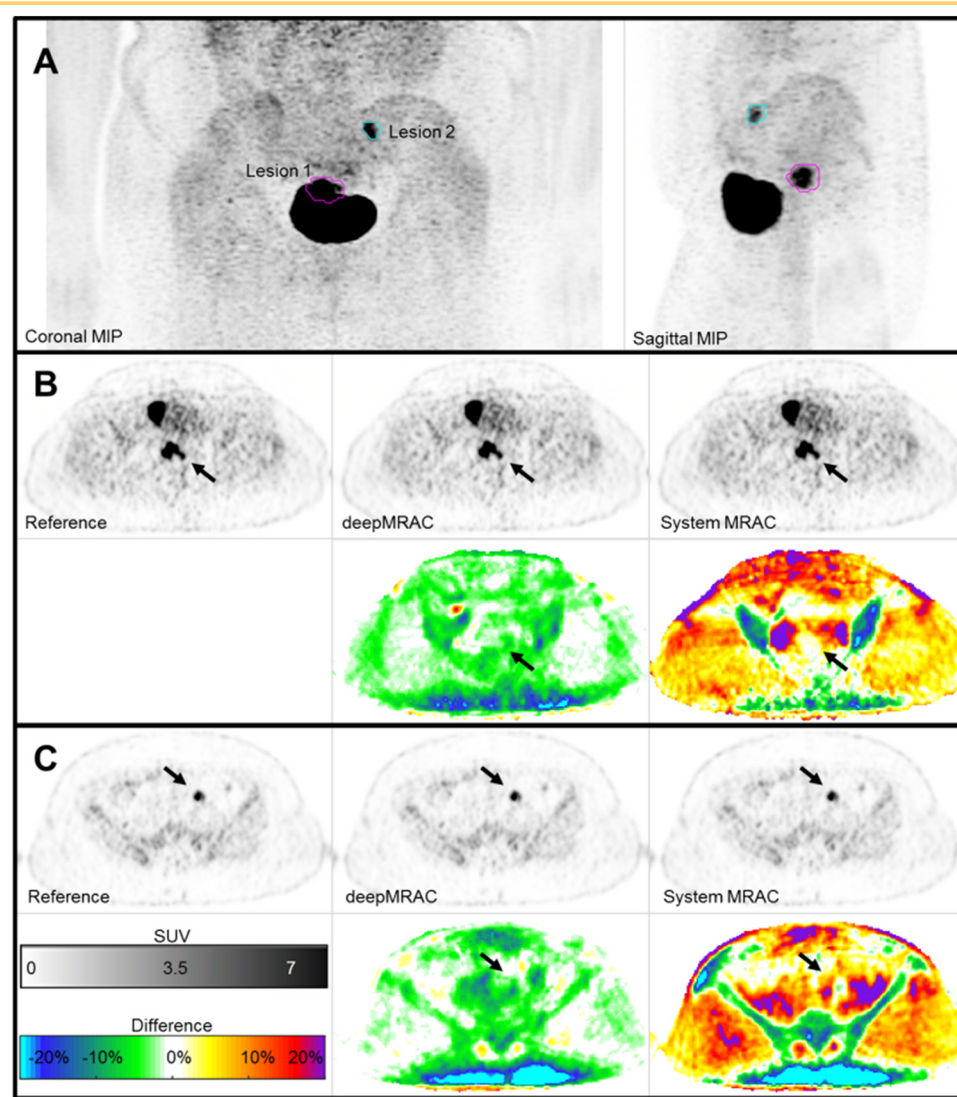
Figure 7. Reconstructed PET images and error maps for 2 example lesions from an example subject. Coronal and sagittal maximum intensity projections (MIPs) showing the location of the 2 lesions (A). Axial PET images reconstructed with CT$_{ref}$, deepMRAC, and the system MRAC, together with respective pixel-wise error maps (B). Arrows indicate the location of the first lesion. PET images and error maps for the second lesion (C).

A primary benefit of deepMRAC is that clinically relevant MRI images can be used as input to the model. This eliminates the need for dedicated MRAC acquisitions such as UTE and ZTE that have limited or no diagnostic value and yet can substantially lengthen scan times in multibed acquisitions. For example, a 2-minute ZTE sequence acquired over 6 bed positions (typical for whole-body PET/MR) would add 12 minutes to the overall scan time (27). Furthermore, more conventional MR acquisition sequences are likely to be more compatible between vendors and different scanner types as proprietary algorithms for postprocessing of Dixon sequences and computationally expensive regridding algorithms necessary for non-Cartesian sequences like ZTE would no longer be acquired. Although we used clinically acquired T2 and LAVA Flex images as input to our model, it is likely that other diagnostic sequences could be equally as effective in such a framework. T2 and LAVA T1 images were selected for our model because they were retrospectively available in all subjects and had a sufficiently large FOV. However, future studies are likely necessary to determine which MR acquisitions are optimal for use in deep learning–based MRAC.

Our deepMRAC method used discretized tissue classes instead of continuous voxel-wise HU estimation. Although we have found continuous mapping effective in MRAC for the brain (32), our initial attempts at training a network for continuous HU prediction in the pelvis resulted in poor performance (not shown). This may be because of the greater complexity in learning a continuous mapping function, or perhaps because of our limited data set. It may also be a limitation of using diagnostic MR images as inputs where bone and air have similar signal intensity, unlike in dedicated ZTE or UTE sequences (19). Whatever the cause, we found networks were less prone to error when learning discrete tissue classes; therefore, we used discrete MRACs in this study. Of course, a limitation of using discrete tissue classes is the assumption of uniform electron density in each tissue class, both within and across subjects. This approximation is clearly not factual, especially in bone, but may be sufficiently accurate for use in MRAC given that our SUV errors were <4% and our PET RMSE$_{discrete}$ was 4.4% when using CT$_{ref-discrete}$ for attenuation correction instead of CT$_{ref}$. And while we determined that assigning cortical bone an HU of 300 minimized the PET errors in our data, different models or body
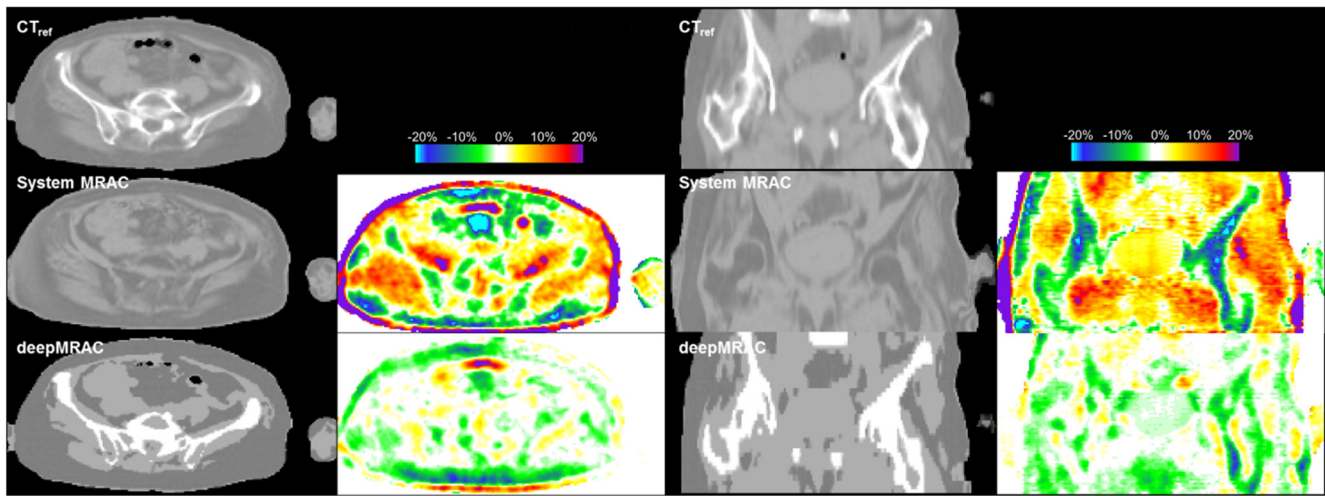
Figure 8. Comparisons of the reference CT (CT$_{ref}$, *top*) to the system MRAC's substitute CT (*middle*) and the deepMRAC substitute CT (*bottom*) are shown for an example subject. Both axial section and coronal section are shown. Resulting pixel-wise PET error maps, relative to the PET image reconstructed with CT$_{ref}$, are also shown.

regions may require different HU assignments. In future work, we intend to further explore networks that produce continuously valued outputs, such as we used in a previous study.

A primary limitation in our study is few subjects. On the other hand, this may also be considered a strength of such an approach: surprisingly, few training subjects were needed to achieve impressive results, a phenomenon commonly reported when using deep learning for medical imaging applications ([18], [19], [33]). It is always uncertain how many training subjects are needed for a given deep learning application, but unlike natural image applications, intersubject variability is often low for medical imaging applications and impressive results can be realized with few training subjects.

A further limitation of our method is the semiautomatic method we used to define bowel gas in the CT$_{ref}$ images. We observed that pockets of bowel gas had a range of intensity values (between −400 to −1000 HU) and sizes, which made it challenging to assign the various air pockets to the appropriate class of tissue. Because of this uncertainty, we segmented only the largest and lowest-intensity air pockets. Owing to different degrees of air–tissue contrasts in different subjects (and even different bed positions), our bowel gas segmentations likely suffered from inconsistencies and hence the comparatively low DSC achieved for bowel gas. Previous PET/MR studies have generally ignored the impact of bowel gas on attenuation correction because of these difficulties (also because their methods relied on image registration for ground truth), which led some to replace gas with tissue-equivalent CT

numbers ([19]). Given that the volume of gas present in the abdomen and pelvis can be sizeable, we aimed to include bowel gas. For example, in the 6 test subjects of this study, we found that the volume of bowel gas in the lower abdomen and pelvis (extending from the bottom of the liver to the bottom of the pelvis) was between 5% and 12% (50–130 mL) of the volume of bone in the same region, where we defined air as ≤200 HU and bone as >200 HU. In our sensitivity tests, we showed that while ignoring bowel gas did not create large systematic PET errors (5.9% RMSE), its effect was actually greater than using a discrete CTAC instead of a continuously-valued CTAC (4.4% RMSE$_{gas}$). In future work, we plan to designate a separate tissue class (or classes) for bowel gas and assess if different MRI sequences can provide better and more consistent contrast for detecting air cavities. In addition, future work should include a direct comparison of our method with that of a UTE/ZTE method and validation of our method using an independent cohort at a separate institution, both of which were not possible with this data set.

CONCLUSION
We have shown that deep learning–based MRAC in the pelvis using only diagnostic MRI sequences is feasible and improves upon the current commercial solution.

Supplemental Materials
Supplemental Material: http://dx.doi.org/10.18383/j.tom.2018.00016.sup.01

REFERENCES

1. Mehranian A, Arabi H, Zaidi H. Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities. Med Phys. 2016;43:1130–1155.

2. Hofmann M, Pichler B, Scholkopf B, Beyer T. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. Eur J Nucl Med Mol Imaging. 2009;36 *Suppl* 1:S93–S104.

3. Zaidi H, Ojha N, Morich M, Griesmer J, Hu Z, Maniawski P, Ratib O, Izquierdo-Garcia D, Fayad ZA, Shao L. Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system. Phys Med Biol. 2011;56:3091–3106.

4. Delso G, Furst S, Jakoby B, Ladebeck R, Ganter C, Nekolla SG, Schwaiger M, Ziegler SI. Performance measurements of the Siemens mMR integrated whole-body PET/MR scanner. J Nucl Med. 2011;52:1914–1922.

5. Martinez-Möller A, Souvatzoglou M, Delso G, Bundschuh RA, Chefd'hotel C, Ziegler SI, Navab N, Schwaiger M, Nekolla SG. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. J Nucl Med. 2009;50:520–526.

6. Spick C, Herrmann K, Czernin J. 18F-FDG PET/CT and PET/MRI perform equally well in cancer: evidence from studies on more than 2,300 patients. J Nucl Med. 2016;57:420–430.

7. Samarin A, Burger C, Wollenweber SD, Crook DW, Burger IA, Schmid DT, von Schulthess GK, Kuhn FP. PET/MR imaging of bone lesions–implications for PET quantification from imperfect attenuation correction. Eur J Nucl Med Mol Imaging. 2012;39:1154–1160.

8. Ladefoged CN, Law I, Anazodo U, St. Lawrence K, Izquierdo-Garcia D, Catana C, Burgos N, Cardoso MJ, Ourselin S, Hutton B, Mérida I, Costes N, Hammers A, Benoit D, Holm S, Juttukonda M, An H, Cabello J, Lukas M, Nekolla S, Ziegler S, Fenchel M, Jakoby B, Casey ME, Benzinger T, Højgaard L, Hansen AE, Andersen FL. A multi-centre evaluation of eleven clinically feasible brain PET/MRI attenuation correction techniques using a large cohort of patients. Neuroimage. 2017; 147:346–359.

9. Paulus DH, Quick HH, Geppert C, Fenchel M, Zhan Y, Hermosillo G, Faul D, Boada F, Friedman KP, Koesters T. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based MR attenuation correction method including bone. J Nucl Med. 2015;56:1061–1066.

10. Arabi H, Zaidi H. Whole-body bone segmentation from MRI for PET/MRI attenuation correction using shape-based averaging. Med Phys. 2016;43:5848–5861.

11. Bezrukov I, Schmidt H, Gatidis S, Mantlik F, Schäfer JF, Schwenzer N, Pichler BJ. Quantitative evaluation of segmentation- and atlas-based attenuation correction for PET/MR on pediatric patients. J Nucl Med. 2015;56:1067–1074.

12. Bini J, Eldib M, Robson PM, Calcagno C, Fayad ZA. Simultaneous carotid PET/MR: feasibility and improvement of magnetic resonance-based attenuation correction. Int J Cardiovasc Imaging. 2016;32:61–71.

13. Mehranian A, Zaidi H. Clinical assessment of emission- and segmentation-based MR-guided attenuation correction in whole-body time-of-flight PET/MR imaging. J Nucl Med. 2015;56:877–883.

14. Salomon A, Goedicke A, Schweizer B, Aach T, Schulz V. Simultaneous reconstruction of activity and attenuation for PET/MR. IEEE Trans Med Imaging. 2011; 30:804–813.

15. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–444.

16. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16:933–951.

17. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging–based attenuation correction for PET/MR imaging. 2018;286:676–684.

18. Jang H, Liu F, Bradshaw TJ, McMillan AB. Technical note: deep learning based MRAC using rapid ultra-short echo time imaging. Med Phys. 2018. [Epub ahead of print]

19. Leynes AP, Yang J, Wiesinger F, Kaushik SS, Shanbhag DD, Seo Y, Hope TA, Larson PEZ. Direct pseudoCT generation for pelvis PET/MRI attenuation correction using deep convolutional neural networks with multi-parametric MRI: zero echo-time and Dixon Deep pseudoCT (ZeDD-CT). J Nucl Med. 2018;59:852–858.

20. Jang H, Liu F, Bradshaw T, McMillan AB. Rapid dual-echo ramped hybrid encoding MR-based attenuation correction (dRHE-MRAC) for PET/MR. Magn Reson Med. 2018;79:2912–2922.

21. Stephen Y, Timothy P, Robert J. Development and evaluation of an articulated registration algorithm for human skeleton registration. Phys Med Biol. 2014;59:1485.

22. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36:61–78.

23. Menze BH, Jakab AB, S., Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Leemput KV. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2015;34:1993–2024.

24. Maier O, Menze BH, von der Gablentz J, Häni L, Heinrich MP, Liebrand M, Winzeck S, Basit A, Bentley P, Chen L, Christiaens D, Dutil F, Egger K, Feng C, Glocker B, Götz M, Haeck T, Halme H-L, Havaei M, Iftekharuddin KM, Jodoin P-M, Kamnitsas K, Kellner E, Korvenoja A, Larochelle H, Ledig C, Lee J-H, Maes F, Mahmood Q, Maier-Hein KH, McKinley R, Muschelli J, Pal C, Pei L, Rangarajan JR, Reza SMS, Robben D, Rueckert D, Salli E, Suetens P, Wang C-W, Wilms M, Kirschke JS, Krämer UM, Münte TF, Schramm P, Wiest R, Handels H, Reyes M. ISLES 2015 – A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med Image Anal. 2017;35: 250–269.

25. Kamnitsas K, Bai B, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, Glocker B, editors. Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation: International Conference on Medical Image Computing for Computer Assisted Intervention (MICCAI); 2017 Sept 14; Quebec, CA.

26. Ouyang J, Chun SY, Petibon Y, Bonab AA, Alpert N, Fakhri GE. Bias atlases for segmentation-based PET attenuation correction using PET-CT and MR. IEEE Trans Nucl Sci. 2013;60:3373–3382.

27. Leynes AP, Yang J, Shanbhag DD, Kaushik SS, Seo Y, Hope TA, Wiesinger F, Larson PEZ. Hybrid ZTE/Dixon MR-based attenuation correction for quantitative uptake estimation of pelvic lesions in PET/MRI. Med Phys. 2017;44:902–913.

28. Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, Wratten C, Martin J, Arm J, Best L, Chandra SS, Fripp J, Menk FW, Greer PB. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. Int J Radiat Oncol Biol Phys. 2015;93:1144–1153.

29. Korhonen J, Kapanen M, Keyriläinen J, Seppälä T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. Med Phys. 2014; 41:011704.

30. Kim J, Glide-Hurst C, Doemer A, Wen N, Movsas B, Chetty IJ. Implementation of a novel algorithm for generating synthetic CT images from magnetic resonance imaging data sets for prostate cancer radiation therapy. Int J Radiat Oncol Biol Phys. 2015;91:39–47.

31. Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, Shen D. Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Trans Med Imaging. 2016;35:174–183.

32. Liu F, Jang H, Kijowski R, Zhao G, Bradshaw TJ, McMillan AB. A deep learning approach for 18F-FDG PET attenuation correction. Eur J Nucl Med Mol Imaging Physics. 2018; In press.

33. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, Heng P-A. 3D deeply supervised network for automated segmentation of volumetric medical images. Med Image Anal. 2017;41:40–54.