Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning

Donghwi Hwang*[1,2], Kyeong Yun Kim*[1,2], Seung Kwan Kang[1,2], Seongho Seo[3], Jin Chul Paeng[2,4], Dong Soo Lee[2,4,5], and Jae Sung Lee[1,2,4,*]

[1]Department of Biomedical Sciences, Seoul National University, Seoul, Korea

[2]Department of Nuclear Medicine, Seoul National University, Seoul, Korea

[3]Department of Neuroscience, College of Medicine, Gachon University, Incheon, Korea

[4]Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Korea

[5]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Suwon, Korea

*Contributed equally for this work.

Name and address for correspondence and reprint requests:

Jae Sung Lee, PhD

Department of Nuclear Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Phone: +82-2-2072-2938; Fax: +82-2-745-2938; E-mail: jaes@snu.ac.kr

Name and address of the first authors:

Donghwi Hwang, PhD Candidate (Currently in training)

Department of Biomedical Sciences, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Phone: +82-2-740-8924; E-mail: hwangdh0310@gmail.com

Kyeong Yun Kim, PhD Candidate (Currently in training)

Department of Biomedical Sciences, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Phone: +82-2-740-8924; E-mail: kykim0126@gmail.com

Running title: Deep Learning

ABSTRACT

Simultaneous reconstruction of activity and attenuation using the maximum likelihood reconstruction of activity and attenuation (MLAA) augmented by time-of-flight (TOF) information is a promising method for positron emission tomography (PET) attenuation correction. However, it still suffers from several problems, including crosstalk artifacts, slow convergence speed, and noisy attenuation maps (μ-maps). In this work, we developed deep convolutional neural networks (CNNs) to overcome these MLAA limitations, and we verified their feasibility using a clinical brain PET data set. **Methods:** We applied the proposed method to one of the most challenging PET cases for simultaneous image reconstruction (^{18}F-FP-CIT PET scans with highly specific binding to striatum of the brain). Three different CNN architectures (convolutional autoencoder (CAE), U-net, hybrid of CAE and U-net) were designed and trained to learn x-ray computed tomography (CT) derived μ-map (μ-CT) from the MLAA-generated activity distribution and μ-map (μ-MLAA). PET/CT data of 40 patients with suspected Parkinson's disease were employed for five-fold cross-validation. For the training of CNNs, 800,000 transverse PET slices and CTs augmented from 32 patient data sets were used. The similarity to μ-CT of the CNN-generated μ-maps (μ-CAE, μ-Unet, and μ-Hybrid) and μ-MLAA was compared using Dice similarity coefficients. In addition, we compared the activity concentration of specific (striatum) and non-specific binding regions (cerebellum and occipital cortex) and the binding ratios in the striatum in the PET activity images reconstructed using those μ-maps. **Results:** The CNNs generated less noisy and more uniform μ-maps than original μ-MLAA. Moreover, the air cavities and bones were better resolved in the proposed CNN outputs. In addition, the proposed deep learning approach was useful for mitigating the crosstalk problem in the MLAA reconstruction. The hybrid network of CAE and U-net yielded the most similar μ-maps to μ-CT (Dice similarity coefficient in the whole head = 0.79 in the bone and 0.72 in air cavities), resulting in only approximately 5% errors in activity and biding ratio quantification. **Conclusion:** The proposed deep learning approach is promising for accurate attenuation correction of activity distribution in TOF PET systems.

The attenuation correction (AC) of annihilation photons is a critical procedure in PET image generation for providing accurate quantitative information on the radiotracer distribution. In current PET/CT systems, the linear attenuation coefficient (μ) for 511 keV photons is converted from the CT Hounsfield unit (1,2). In PET / magnetic resonance imagery (MRI), various approaches, including DIXON and UTE MRI segmentation- and atlas-based algorithms, have been suggested (3-6). However, a limitation of CT-based PET AC is the artifacts attributed to the position mismatch between the PET and CT (7-9). MRI-based PET AC remains far from ideal on account of the inaccurately estimated linear attenuation coefficients (μ-values) in the skeletal structures and heterogeneous soft tissues (10-12). In particular, bones are poorly identified in whole-body PET/MRI studies (13) and local MRI signal loss produced by metallic implants results in the considerable error in image segmentation.

Simultaneous reconstruction of activity and attenuation using only emission data is a promising method for the PET AC augmented by the recent advancement of TOF technology (14-17). Because no anatomical images are necessary for the AC if the simultaneous reconstruction works properly, it is a potentially significant approach to overcoming the above-mentioned limitations of PET AC in PET/CT and PET/MRI (18-20). Among the simultaneous reconstruction algorithms for PET AC, the MLAA method has the advantages of providing an μ-map and enabling the incorporation of prior knowledge of the μ-values for global scaling (15,16,21,22). However, because of the limited timing resolution of current clinical PET scanners, the MLAA suffers from several problems, including the crosstalk artifacts (between activity and attenuation maps), slow convergence speed, and noisy attenuation maps (23,24).

Recently, deep learning has outperformed the traditional machine learning and Bayesian approaches in many different applications (25,26). In addition, recent studies have shown the remarkable advancements in the noise reduction of x-ray CT based on deep learning technology (27,28). Accordingly, it is of interest whether the deep learning approach can mitigate the limitations of MLAA simultaneous reconstruction. In this study, we therefore designed deep CNNs to be suitable for MLAA output (activity distribution and μ-map) processing. We examined the quality improvement of MLAA μ-maps and emission PET images by applying the deep

learning with the focus on noise and crosstalk reduction.

We applied this new approach to one of the most challenging clinical PET cases for simultaneous reconstruction (brain dopamine transporter imaging). The crosstalk between the activity and attenuation is severe and the background noise level is high in the dopaminergic PET images because of the highly specific binding of the tracers only in the striatum of brain.

MATERIALS AND METHODS

Data Set

The ^{18}F-fluorinated-N-3-fluoropropyl-2-β-carboxymethoxy-3-β-(4-iodophenyl)nortropane (^{18}F-FP-CIT) brain PET/CT scan data of 40 patients (16 males and 24 females, age = 67.5 ± 9.2 years) with suspected Parkinson's disease were retrospectively analyzed. In 14 subjects of 40, the tracer uptake in both basal ganglia was preserved. The retrospective use of the scan data and waiver of consent were approved by the Institutional Review Board of our institute. The PET/CT data were acquired using a Siemens Biograph mCT40 scanner (Siemens Healthcare, Knoxville, TN). The PET scanner achieves the effective timing resolution of 580 ps. The PET/CT imaging was obtained for 10 min in a single PET bed position 90 min after the intravenous injection of ^{18}F-FP-CIT (189.7 MBq on average). The head of each participant was positioned in a head holder attached to the patient bed, and the PET/CT scan followed the routine clinical protocol for brain studies (topogram, CT, and emission PET scans). The CT images were reconstructed in a 512 × 512 × 149 matrix with voxel sizes of 0.59 × 0.59 × 1.5 mm and converted into the μ-map for 511 keV photons (μ-CT, 200 × 200 × 109; 2.04 × 2.04 × 2.03 mm).

We reconstructed all data sets using the ordered subset expectation maximization (OSEM) with μ-CT (3 iterations, 21 subsets, 5 mm Gaussian post-filter) and MLAA with the TOF information (8 iterations and 21 subsets, 5 mm Gaussian post-filter) into 200 × 200 × 109 matrices. To correct the global scaling problem, the boundary constraint suggested in the original TOF MLAA paper (*15*) was applied during the attenuation image estimation in the MLAA.

To evaluate the performance of proposed CNNs, we performed five-fold cross-validation. The 40 patient data sets were randomly partitioned into five groups (eight in each group). The CNNs were trained with four groups and tested with the other one. This cross-validation process with different test set was repeated five times. For the CNN training and testing, the activity distribution and μ-map derived from the MLAA (λ-MLAA and μ-MLAA) were used as input *X*, and μ-CT was used as output *Y*. All the input and output images were used in two-dimensional (2D) slice format.

Network Architecture

We tested three different CNN architectures (**Fig. 1**). The first one was the CAE. The autoencoder was originally proposed for unsupervised feature learning; nevertheless, it also showed good performance for image restoration and denoising networks (*29*). The second one was U-net, which showed excellent performance in various tasks, including image segmentation and denoising (*30*). U-net structures are similar to those of the CAE. However, unlike CAE, U-net supplements the contracting path that enables high-resolution features to be combined in the output layers. The third one was the hybrid form of CAE and U-net, which we propose herein to prevent the noise propagation from the high-frequency feature of the PET activity distribution (hybrid network).

The three networks (CAE, U-net, and hybrid) consisted of convolution layers, rectified linear units (an activation function defined as $f(x)=\max(0, x)$ and employed to provide non-linearity in learning model), 2×2 max-pooling layers, deconvolution layers, and a 1×1 convolution layer (**Fig. 1**). The max-pooling, a reduction operation for calculating the maximum value in each rectangular window, was required to reduce the number of parameters of the network and provide shift invariant characteristic to CNN. In the first layer, we performed a convolution with a $3 \times 3 \times 2$ kernel to merge two input data sets (MLAA activity distribution and μ-map). Each convolution and deconvolution layer except the first convolution layer was composed of a 3×3 kernel and rectified linear units. The last 1×1 convolution layer performed a role in scaling. Number of feature maps in the first layer was empirically determined to yield the best results. We implemented the networks using the TensorFlow, an open source library for deep learning (*31*).

Data Augmentation and Training

Because the number of parameters in the deep networks was too large to be estimated from the limited patient data set, we had to increase the training data (data augmentation). We conducted the data augmentation by rotating the images by {−6°, −3°, 0°, 3°, 6°} in three-dimensional (3D) orthogonal planes (5 × 5 × 5 = 125 times of augmentation). Additionally, we used images flipped in the transverse plane to double the training set. Accordingly, the total number of slices available for training was 32 (patient) × 109 (transverse plane) × 125 (rotating) × 2 (flip) = 872,000. Among them, only 800,000 slices were used as the training set after eliminating the slices with only the negligibly small pixel values and the last five slices at the bottom.

The cost function was the L2-norm between the MLAA- and CT-derived μ-maps. The cost function was minimized using the adaptive moment estimation method (*32*). Weights in the networks were initialized using the Xavier initialization method, which engendered a faster convergence rate compared to uniform or Gaussian random initialization (*33*). To prevent network overfitting, a part of nodes was dropped out (*34*). In each convolution layer, the dropout probability (ratio of remaining node number out of total) was 0.7. The batch size was 60, and the number of epochs (the number of times the algorithm sees the entire data set) was 6. When using the Ryzen 1700X CPU with a GTX 1080 GPU, each epoch involved approximately 300 min.

Image Analysis

The μ-maps obtained using the MLAA before (μ-MLAA) and after (μ-CAE, μ-Unet, and μ-Hybrid) applying the deep CNNs to the test set were compared to the μ-CT, the ground truth. The similarity of μ-maps was evaluated using the Dice similarity coefficient (*3,35*), which measured the overlap of the segmented bone and air regions according to the following equations:

$$D = \frac{2 \times N_{(\mu-CT \cap \mu-PET)}}{N_{\mu-CT} + N_{\mu-PET}}$$

where $N_{\mu-CT}$ and $N_{\mu-PET}$ are respectively the number of bone (or air) voxels in the μ-maps derived from CT and PET (emission only) data (3,35). $N_{(\mu-CT \cap \mu-PET)}$ indicates the number of overlapped voxels between CT and PET μ-maps. In the μ-maps, the voxels having a μ-value greater than 0.1134 (= 300 Hounsfield units) were classified as bone; those having a μ-value less than 0.0475 (= -500 Hounsfield units) were denoted as air. Additionally, the voxels having a μ-value between them were regarded as soft tissue (3,36).

For the comparison with the ground truths of the PET activity distribution obtained using OSEM reconstruction with μ-CT, the activity images were generated using the same OSEM algorithm and parameters (8 iterations and 21 subsets, 5 mm Gaussian post-filter) with μ-MLAA, μ-CAE, μ-Unet, and μ-Hybrid (**Fig. 2**). The ground truth PET activity was spatially normalized using an in-house ^{18}F-FP-CIT PET template and Statistical Parametric Mapping 8 (SPM8) software (http://www.fil.ion.ucl.ac.uk/spm). The same transformation parameters were applied to the others. Then, we measured the PET activity concentration in four regions of interest (head of caudate nucleus, putamen, occipital cortex, and cerebellum) using an automatic region of interest delineation method with statistical probabilistic anatomic maps (37,38). For the comparison, the relative ratio of specific binding (BR = [$C_{specific} - C_{nonspecific}$] / $C_{nonspecific}$) was calculated ($C_{specific}$ and $C_{nonspecific}$: activity concentrations in specific and nonspecific (cerebellum or occipital cortex) binding regions) (3).

RESULTS

The CNNs remarkably reduced the noise and crosstalk in MLAA-generated μ-maps (μ-MLAA). In **Figure 3**, the CNN-generated μ-maps (μ-CAE, μ-Unet, and μ-Hybrid) of a patient are compared to the μ-CT and μ-MLAA. As expected, the CNNs generated less noisy and more uniform images than μ-MLAA. Among the CNNs, the CAE and U-net yielded the most blurred and sharpened μ-maps, respectively. The air cavities and bones were better resolved in the proposed CNN outputs compared to the μ-MLAA. However, the details of these structures did not perfectly match with the μ-CT. Moreover, slight discontinuities of air cavities and bone structures still appeared in the sagittal and coronal planes owing to the application of the CNNs to the 2D slices. **Figure 3** also shows that the proposed deep learning approach is useful for mitigating the crosstalk problem in MLAA reconstruction. The red arrows on μ-MLAA point to the striatal region where the crosstalk between activity and attenuation is substantial in MLAA outputs. This artifact disappears in the μ-maps corrected by deep learning (μ-CAE, μ-Unet, and μ-Hybrid).

The CNN-generated μ-maps showed higher similarity with the μ-CT than the original μ-MLAA did. In **Figure 4**, root-mean square errors from the μ-CT are plotted across the slice axial location (average of all 40 test subjects)**.** The CNN-generated μ-maps yields fewer errors than μ-MLAA in almost all axial locations. The hybrid network outperformed the CAE and U-net at the top of the head, and it achieved approximately 50% error reduction relative to the original MLAA in the μ-value estimation. The bias and root-mean square error of μ-maps relative to the μ-CT are summarized in **Supplemental Figure 1**.

As shown in **Table 1**, the Dice similarity coefficients measured in the whole head and only the cranial bone regions for air and bone are generally much higher in the CNN-generated μ-maps. The standard deviations of the Dice similarity coefficients are considerably smaller in the CNN-generated μ-maps than in those of μ-MLAA, indicating the improvement in the consistency of μ-value estimation. In μ-MLAA, the skull intensity and thickness is under- or over-estimated in some regions (yellow arrows in **Fig. 3**). However, the CNNs properly correct these errors. Among

the CNNs, the hybrid network and CAE yield the respective highest and lowest Dice similarity coefficients in all the regions. The **Supplemental Figure 2** shows that these results are consistent across all the cross-validation sets.

The enhancement in μ-map quality and accuracy by applying the deep CNNs improved the accuracy in the quantification of the regional activity and binding ratio (BR) of ^{18}F-FP-CIT PET. The percentage error map of the spatially normalized activity distribution (average of 40 test subjects) is compared in **Figure 5**, indicating the reduced error in activity distribution with U-net and hybrid network. Meanwhile, **Figure 6** shows the percent error in (A) activity and (B) BR estimation relative to the ground truth (OSEM with μ-CT). The μ-MLAA yields a negative bias in activity quantification that is higher than 10% in the occipital cortex and striatum. The error is reduced using μ-Unet, and μ-Hybrid.

DISCUSSION

Supervised and unsupervised machine learning methods based on artificial neural networks have been investigated for various biomedical engineering applications (*39,40*). Learning the difference between the patient and control data, and predicting the prognosis after treatment based on region of interest-driven features, was the main application field in nuclear medicine image interpretation and processing (*41-43*). Additionally, data-driven blind source separation techniques based on unsupervised neural network were successfully applied to dynamic PET data for the separation of various physiological and anatomical components (*44-47*). The use of artificial neural network techniques has been also suggested for more accurate and reliable determination of the annihilation photon interaction position in PET detector blocks (*48,49*). Meanwhile, deep learning, an emerging technology in machine learning, is showing its initial impact on the medical imaging field (*50*). However, problem-specific design and optimization of deep networks and rigorous validation with real clinical data are required to justify the medical use of this emergent technology.

One of the main limitations of simultaneous activity and attenuation reconstruction is the crosstalk artifacts between the activity and attenuation in output images. These crosstalk artifacts are most severe in regions with a high contrast against the background, which may be the abnormal uptake of radiotracer requiring high accuracy in activity quantification. The deep CNNs proposed in this paper outperformed the original MLAA algorithm in suppressing the crosstalk and noise in the dopamine transporter PET images. The mitigation of crosstalk artifacts by the CNNs was not simply the consequence of reducing the noise (or suppressing the high-frequency features) in μ-map and recognizing the location of crosstalk artifacts (**Supplemental Fig. 4**). Only when the activity and attenuation information were jointly processed by the 3D convolution kernel at the first layer of the networks, this crosstalk was successfully suppressed. The joint feature learning from both the activity and attenuation at the early stage was also useful for the accurate restoration of bone structures and air cavities in μ-maps (**Fig. 3**). The lower radioactivity in the bone and air relative to the soft tissue and cerebrospinal fluid would

enable the CNNs to learn how to correctly differentiate them.

Most CNN parameters were empirically determined through trial and error. The performance of CNN was not much influenced by the kernel size (i.e., five or seven yielded similar results as three did). The 20 feature maps in the first layer yielded the lowest and most stable learning curve. On the contrary, the learning curve did not converge with feature maps smaller than 12 and showed overshoot at early iterations with the maps larger than 28. Learning rate and other parameters were also determined mostly while observing the learning curves.

The CNNs trained in this study yielded better μ-maps than our multiphase level-set based UTE MRI segmentation (*3*) with respect to the similarity with μ-CT. Particularly, the Dice similarity coefficients for air cavities were remarkably higher in the present study (0.72 *versus* 0.61 in the whole head and 0.74 *versus* 0.62 in the cranial region). Although the hybrid method has the higher Dice coefficient than U-net, the U-net performs slightly better than the hybrid method in the activity quantification (**Figs. 5 and 6**). That is because the Dice coefficient measures similarity in terms of segmented region overlap but it does not measure similarity in terms of quantitative values. As shown in **Supplemental Figure 1**, the U-net yielded lower bias in average μ-values than the other methods, explaining why the activity maps estimated by the hybrid method are less accurate than those estimated by the U-net in spite of the better segmentation.

There are some existing works on applying the deep learning to predict CT attenuation map based on T1-weighted MR images or combination of Dixon and ZTE images (*51, 52*). The approach using the Dixon and ZTE images would be more physically relevant than T1 MRI-based approach because the Dixon and ZTE sequences provide more direct information on the tissue composition than T1 sequence. The method proposed in this study has same physical relevance as the DIXON/ZTE approach but does not require the acquisition of additional MR images.

CONCLUSIONS

In this work, we developed deep CNNs to overcome the main limitations of the MLAA simultaneous reconstruction algorithm. We verified their feasibility using an ^{18}F-FP-CIT brain PET data set. The proposed deep learning approach remarkably enhanced the quantitative accuracy of simultaneously estimated MLAA μ-maps by reducing the noise and crosstalk artifacts. The hybrid network of CAE and U-net yielded μ-maps the most similar to μ-CT (Dice similarity coefficient in the whole head = 0.79 in the bone and 0.72 in air cavities), resulting in only approximately 5% errors in activity and binding ratio quantification. Because the proposed method requires no transmission data, anatomical image, or atlas/template for PET attenuation correction, it has potential to replace the conventional attenuation correction methods in stand-alone PET, PET/CT, and PET/MRI.

ACKNOWLEDGMENTS

REFERENCES

1.	Burger C, Goerres G, Schoenes S, Buck A, Lonn A, Von Schulthess G. PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. *Eur J Nucl Med Mol Imaging.* 2002;29:922-927.

2.	Kinahan PE, Hasegawa BH, Beyer T. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Semin Nucl Med*. 2003;33:166-179.

3.	An HJ, Seo S, Kang H, et al. MRI-based attenuation correction for PET/MRI using multiphase level-set method. *J Nucl Med.* 2016;57:587-593.

4.	Keereman V, Fierens Y, Broux T, De Deene Y, Lonneux M, Vandenberghe S. MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. *J Nucl Med.* 2010;51:812-818.

5.	Martinez-Möller A, Souvatzoglou M, Delso G, et al. 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.	Yang J, Jian Y, Jenkins N, et al. Quantitative evaluation of atlas-based attenuation correction for brain PET in an integrated time-of-flight PET/MR imaging system. *Radiology.* 2017:284:169-179.

7.	Gould KL, Pan T, Loghin C, Johnson NP, Guha A, Sdringola S. Frequent diagnostic errors in cardiac PET/CT due to misregistration of CT attenuation and emission PET images: a definitive analysis of causes, consequences, and corrections. *J Nucl Med.* 2007;48:1112-1121.

8.	Liu C, Pierce II LA, Alessio AM, Kinahan PE. The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging. *Phys Med Biol.* 2009;54:7345-7362.

9.	McQuaid SJ, Hutton BF. Sources of attenuation-correction artefacts in cardiac PET/CT and SPECT/CT. *Eur J Nucl Med Mol Imaging.* 2008;35:1117-1123.

10.	Keller SH, Holm S, Hansen AE, et al. Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI. *MAGMA.* 2013;26:173-181.

11.	Kim JH, Lee JS, Song I-C, Lee DS. Comparison of segmentation-based attenuation correction methods for PET/MRI: evaluation of bone and liver standardized uptake value with oncologic PET/CT data. *J Nucl Med.* 2012;53:1878-1882.

12.	Yoo HJ, Lee JS, Lee JM. Integrated whole body MR/PET: where are we? *Korean J Radiol.* 2015;16:32-49.

13.	Fraum TJ, Fowler KJ, McConathy J. Conspicuity of FDG-avid osseous lesions on PET/MRI versus PET/CT: a

quantitative and visual analysis. *Nucl Med Mol Imaging.* 2016;50:228-239.

14.   Defrise M, Rezaei A, Nuyts J. Time-of-flight PET data determine the attenuation sinogram up to a constant. *Phys Med Biol.* 2012;57:885-899.

15.   Rezaei A, Defrise M, Bal G, et al. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. *IEEE Trans Med Imaging.* 2012;31:2224-2233.

16.   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.

17.   Son JW, Kim KY, Yoon HS, et al. Proof-of-concept prototype time-of-flight PET system based on high-quantum-efficiency multi-anode PMTs. *Med Phys.* 2017;44:5314-5324.

18.   Lee JS, Kovalski G, Sharir T, Lee DS. Advances in imaging instrumentation for nuclear cardiology. J *Nucl Cardiol.* 2017:1-14 [Epub ahead of print].

19.   Presotto L, Busnardo E, Perani D, Gianolli L, Gilardi M, Bettinardi V. Simultaneous reconstruction of attenuation and activity in cardiac PET can remove CT misalignment artifacts. *J Nucl Cardiol.* 2016;23:1086-1097.

20.   Rezaei A, Michel C, Casey ME, Nuyts J. Simultaneous reconstruction of the activity image and registration of the CT image in TOF-PET. *Phys Med Biol.* 2016;61:1852-1874.

21.   Boellaard R, Hofman M, Hoekstra O, Lammertsma A. Accurate PET/MR quantification using time of flight MLAA image reconstruction. *Mol Imaging Biol.* 2014;16:469-477.

22.   Nuyts J, Dupont P, Stroobants S, Benninck R, Mortelmans L, Suetens P. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Trans Med Imaging.* 1999;18:393-403*.*

23.   Chun SY, Kim KY, Lee JS, Fessier JA. Joint estimation of activity distribution and attenuation map for TOF-PET using alternating direction method of multiplier. *Proceedings of 13th IEEE International Symposium on Biomedical Imaging*.

24.   Mehranian A, Zaidi H. Joint estimation of activity and attenuation in whole-body TOF PET/MRI using constrained gaussian mixture models. *IEEE Trans Med Imaging.* 2015;34:1808-1821.

25.   Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell.* 2013;35:1798-1828.

26.   Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks.* 2015;61:85-117.

27.　　Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional neural network. *Biomed Opt Express.* 2017;8:679-694.

28.　　Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys.* 2017;44:e360-e375.

29.　　Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res.* 2010;11:3371-3408.

30.　　Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015;234-241.

31.　　Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:160304467.* 2016.

32.　　Kingma D, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980.* 2014.

33.　　Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010.

34.　　Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929-1958.

35.　　Juttukonda MR, Mersereau BG, Chen Y, et al. MR-based attenuation correction for PET/MRI neurological studies with continuous-valued attenuation coefficients for bone through a conversion from R2* to CT-Hounsfield units. *Neuroimage.* 2015;112:160-168.

36.　　Catana C, van der Kouwe A, Benner T, et al. Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype. *J Nucl Med.* 2010;51:1431-1438.

37.　　Kang KW, Lee DS, Cho JH, et al. Quantification of F-18 FDG PET images in temporal lobe epilepsy patients using probabilistic brain atlas. *Neuroimage.* 2001;14:1-6.

38.　　Lee JS, Lee DS. Analysis of functional brain images using population-based probabilistic atlas. *Curr Med Imaging Rev.* 2005;1:81-87.

39.　　Mohammadi MR, Khaleghi A, Nasrabadi AM, Rafieivand S, Begol M, Zarafshan H. EEG classification of ADHD and normal children using non-linear features and neural network. *Biomed Eng Lett.* 2016;6:66-73.

40.   Yoo Y. On predicting epileptic seizures from intracranial electroencephalography. *Biomed Eng Lett.* 2017;7:1-5.

41.   Acton PD, Newberg A. Artificial neural network classifier for the diagnosis of Parkinson's disease using [99mTc]TRODAT-1 and SPECT. *Phys Med Biol.* 2006;51:3057-3066.

42.   Lee JS, Lee DS, Kim S-K, et al. Localization of epileptogenic zones in F-18 FDG brain PET of patients with temporal lobe epilepsy using artificial neural network. *IEEE Trans Med Imaging.* 2000;19:347-355.

43.   Preis O, Blake MA, Scott JA. Neural network evaluation of PET scans of the liver: a potentially useful adjunct in clinical interpretation. *Radiology.* 2011;258:714-721.

44.   Lee JS, Lee DD, Choi S, Park KS, Lee DS. Non-negative matrix factorization of dynamic images in nuclear medicine. *IEEE Nuclear Science Symposium Conference Record*, 2001;2027-2030.

45.   Lee JS, Lee DS, Ahn JY, et al. Blind separation of cardiac components and extraction of input function from h215o dynamic myocardial pet using independent component analysis. *J Nucl Med.* 2001;42:938-943.

46.   Naganawa M, Kimura Y, Ishii K, Oda K, Ishiwata K, Matani A. Extraction of a plasma time-activity curve from dynamic brain PET images based on independent component analysis. *IEEE Trans Biomed Eng.* 2005;52:201-210.

47.   Su K-H, Wu L-C, Liu R-S, Wang S-J, Chen J-C. Quantification method in [18F] fluorodeoxyglucose brain positron emission tomography using independent component analysis. *Nucl Med Commun.* 2005;26:995-1004.

48.   Michaud J-B, Tetrault M-A, Beaudoin J-F, et al. Sensitivity increase through a neural network method for LOR recovery of ICS triple coincidences in high-resolution pixelated-detectors PET scanners. *IEEE Trans Nucl Sci.* 2015;62:82-94.

49.   Wang Y, Zhu W, Cheng X, Li D. 3D position estimation using an artificial neural network for a continuous scintillator PET detector. *Phys Med Biol.* 2013;58:1375-1390.

50.   Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221-248.

51.   Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys.* 2017;44:1408-1419.

52.   Leynes AP, Yang J, Wiesinger F, et al. 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.* 2017:jnumed. 117.198051 [Epub ahead of print].

FIGURE LEGENDS
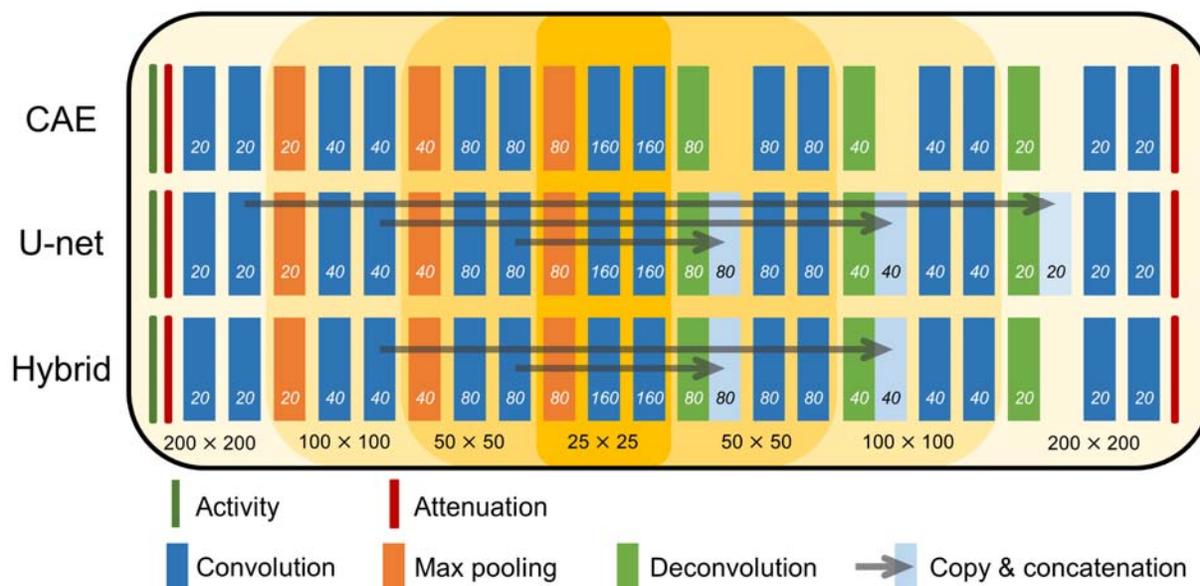
FIGURE 1. CNN architectures used to learn CT-derived μ-map (μ-CT) from the MLAA-derived activity distribution and μ-map (λ-MLAA and μ-MLAA). (A) CAE. (B) U-net. (C) Hybrid network of CAE and U-net. The green and red vertical strips at the far left indicate the inputs to the CNN, and the red stripes on the right indicate the output. Each box represents a multi-channel feature map. The number of feature maps and the dimension of each feature map are denoted on the interior and bottom of the box. The data flow is left to right through the contracting path to capture context and symmetric expanding path to recover the image. Gray arrows stand for copying feature maps and sky blue boxes are copied feature map.
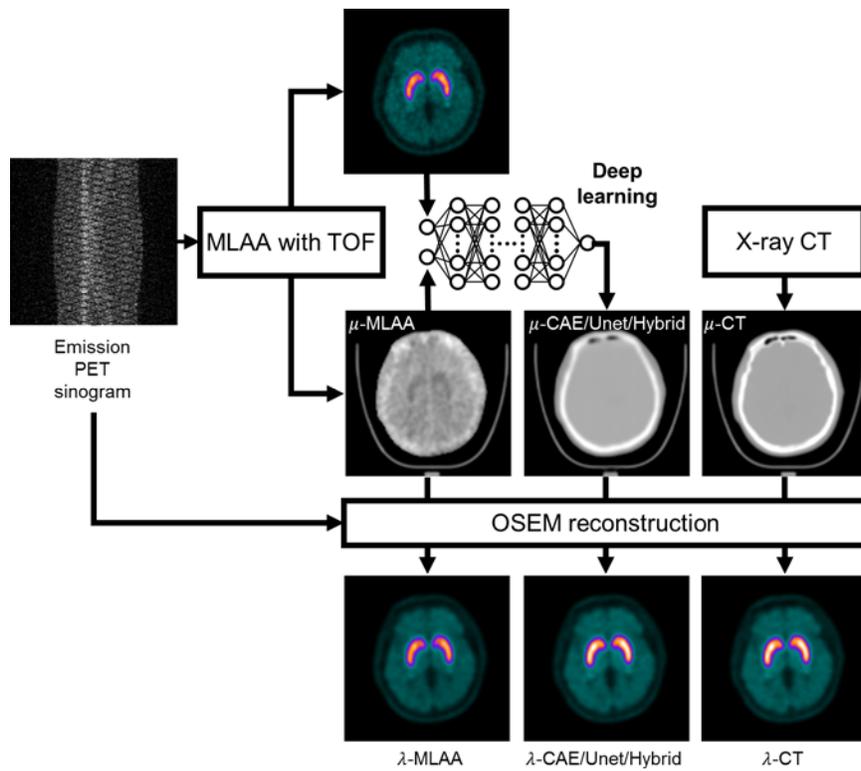
FIGURE 2. Flow chart of image analysis. For the comparison, an emission PET sinogram was reconstructed by employing μ-maps obtained using the MLAA before (μ-MLAA) and after (μ-CAE, μ-Unet, and μ-Hybrid) applying the deep CNNs and ground truth μ-CT.
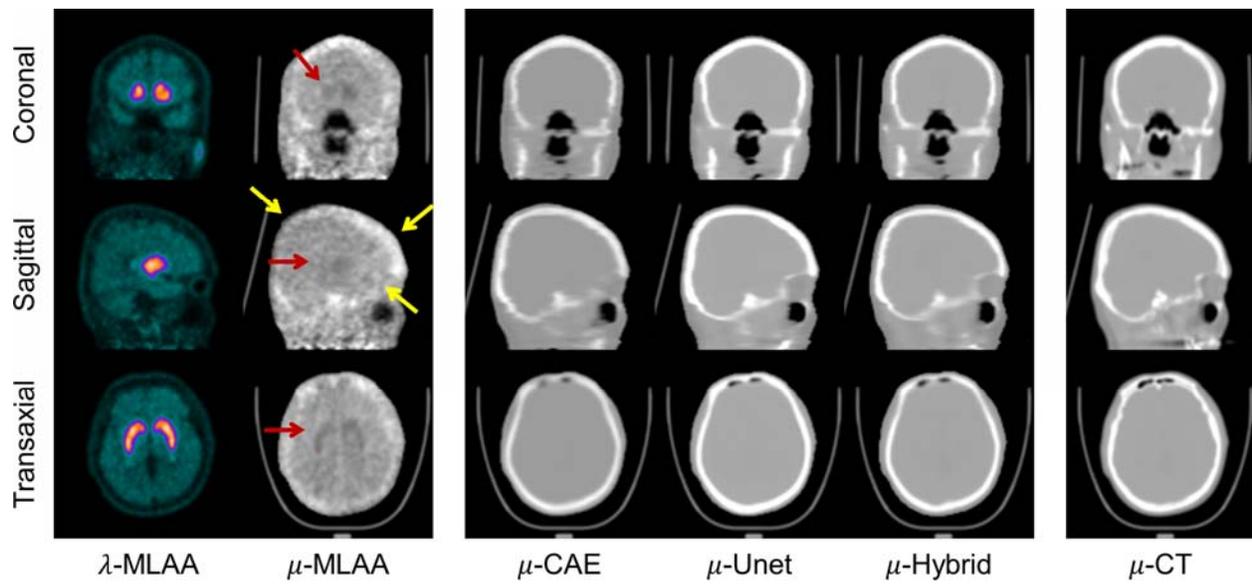
FIGURE 3. Comparison of CNN outputs (μ-CAE, μ-Unet, and μ-Hybrid) to the μ-MLAA and μ-CT. The red and yellow arrows respectively indicate the crosstalk artifacts and bone estimation error shown in μ-MLAA.
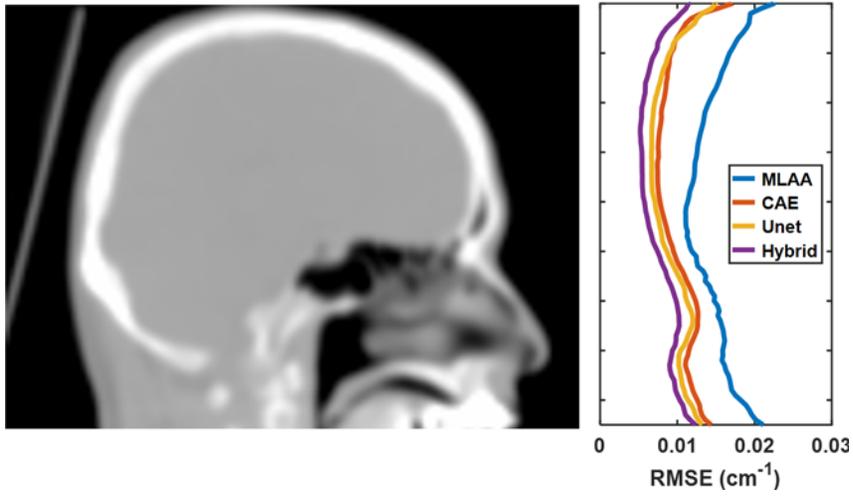
FIGURE 4. Root-mean square errors relative to the μ-CT plotted across the slice axial location (average of 40 test sets).
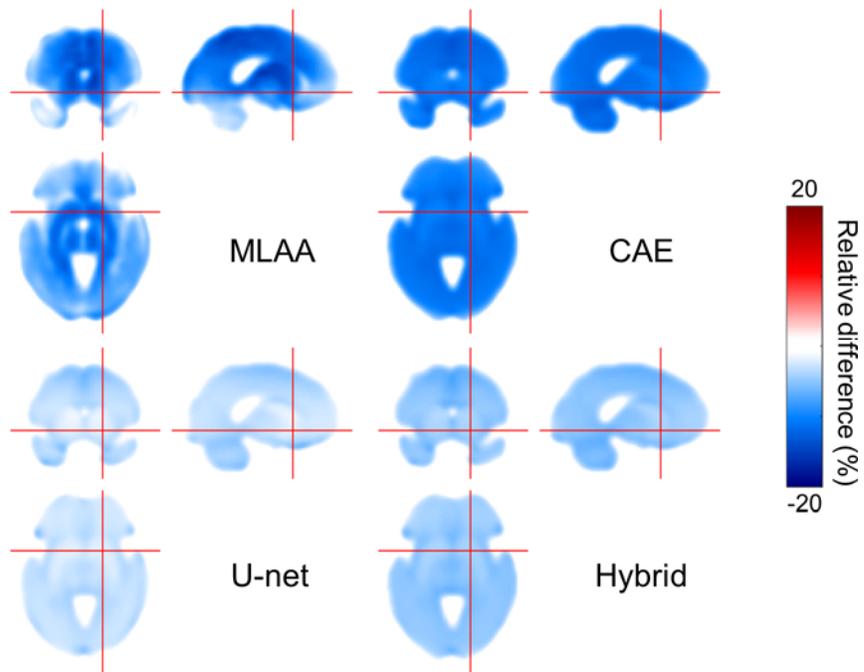
Figure 5. Percentage error map of spatially normalized activity distribution (average of 40 test sets)
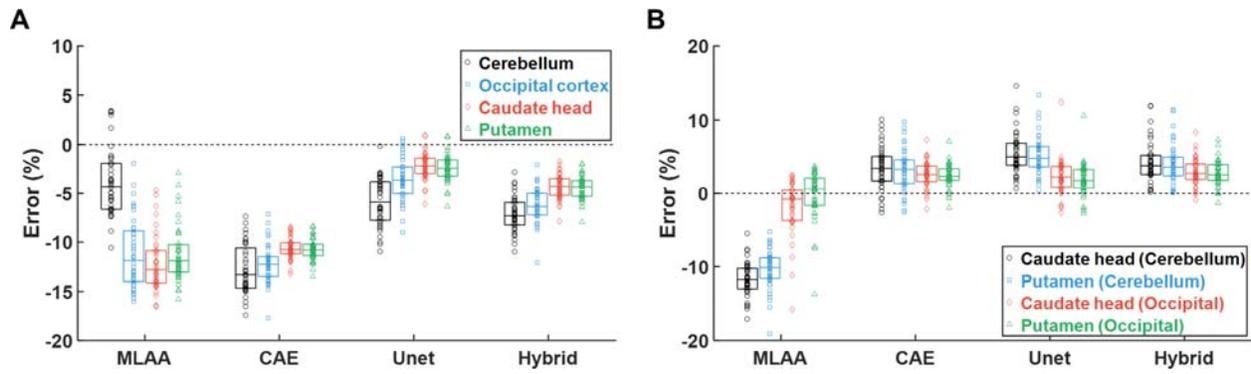
Figure 6. Percent error in (A) activity and (B) binding ratio (BR) estimation relative to the ground truth (OSEM with μ-CT). Each horizontal bar and vertical box indicates the median and standard deviation, respectively. In (B), specific and non-specific regions for the BR calculation are indicated as 'specific (non-specific)'.

TABLE 1

Dice Similarity Coefficients with CT-Derived Attenuation Map for Whole Head and Cranial Bone Region (Mean ± standard deviation). The results of analysis of variation and post-hoc tests are now shown in **Supplemental Figure 3**.

Method	Whole head		Cranial region	
	D_{bone}	D_{air}	D_{bone}	D_{air}
MLAA	0.374± 0.058	0.317 ± 0.070	0.399 ± 0.063	0.426 ± 0.062
CNN (CAE)	0.717 ± 0.047	0.513 ± 0.057	0.747 ± 0.047	0.523 ± 0.063
CNN (U-net)	0.787 ± 0.042	0.575 ± 0.047	0.801 ± 0.043	0.580 ± 0.053
CNN (Hybrid)	0.794 ± 0.037	0.718 ± 0.048	0.810 ± 0.038	0.738 ± 0.044

<Supplemental Figures>

Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning

Donghwi Hwang*[1,2], Kyeong Yun Kim*[1,2], Seung Kwan Kang[1,2], Seongho Seo[3], Jin Chul Paeng[2,4], Dong Soo Lee[2,4,5], and Jae Sung Lee[1,2,4,*]

[1]Department of Biomedical Sciences, Seoul National University, Seoul, Korea

[2]Department of Nuclear Medicine, Seoul National University, Seoul, Korea

[3]Department of Neuroscience, College of Medicine, Gachon University, Incheon, Korea

[4]Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Korea

[5]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Suwon, Korea

***Contributed equally for this work.**

Supplemental Figure 1. Bias and root-mean square error of μ-maps relative to the μ-CT

Hwang D, Kim KY *et al.*, Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning
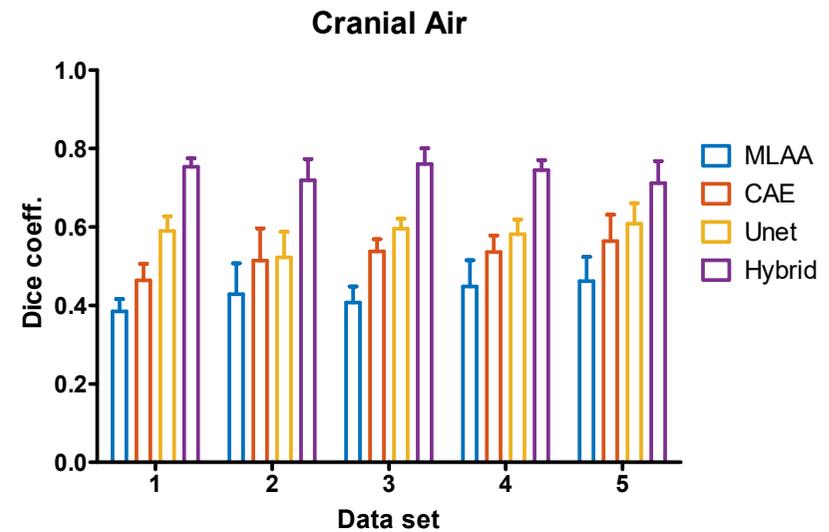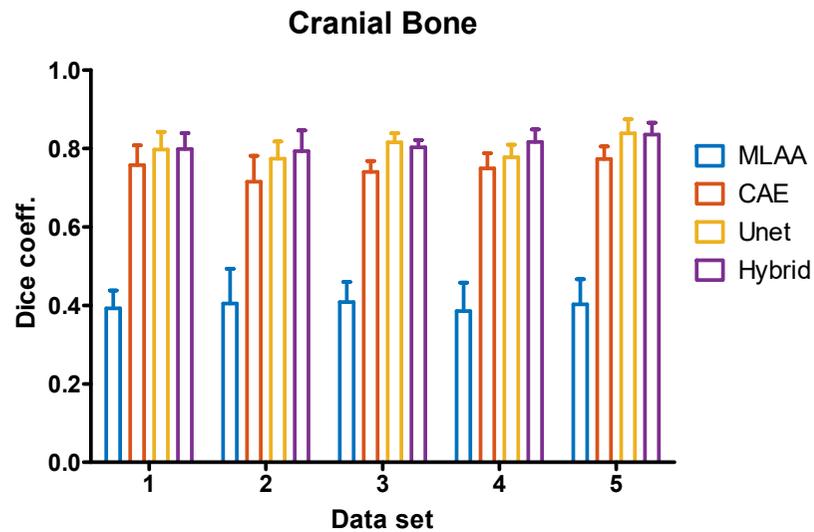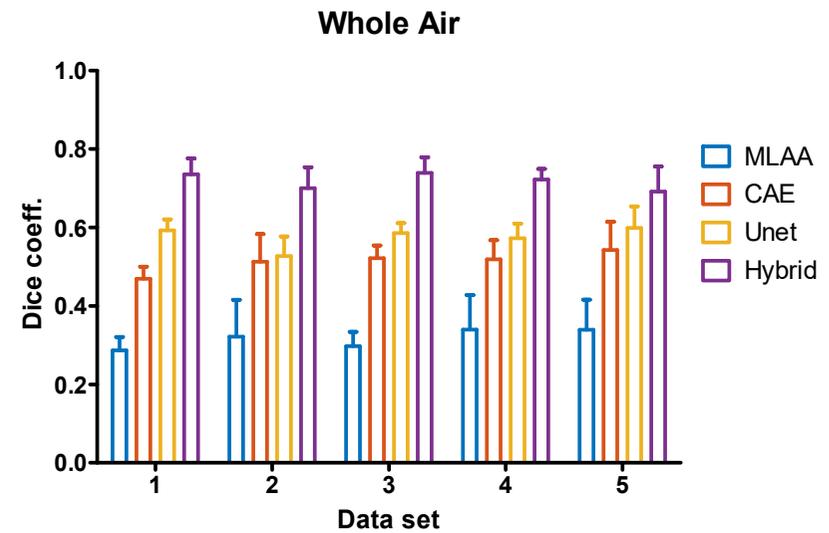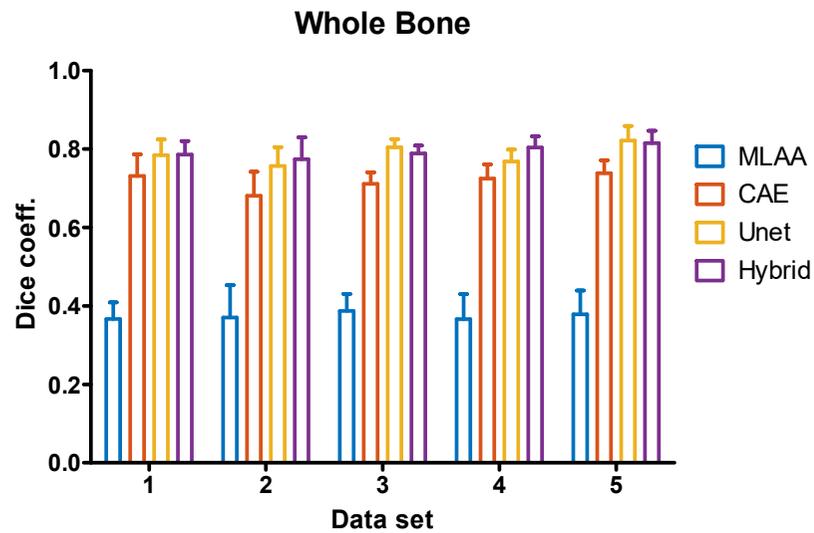
Supplemental Figure 2. The Dice similarity coefficients measured in the whole head and only the cranial bone regions for air and bone (results from 5-fold cross-validations)

Hwang D, Kim KY *et al.*, Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning
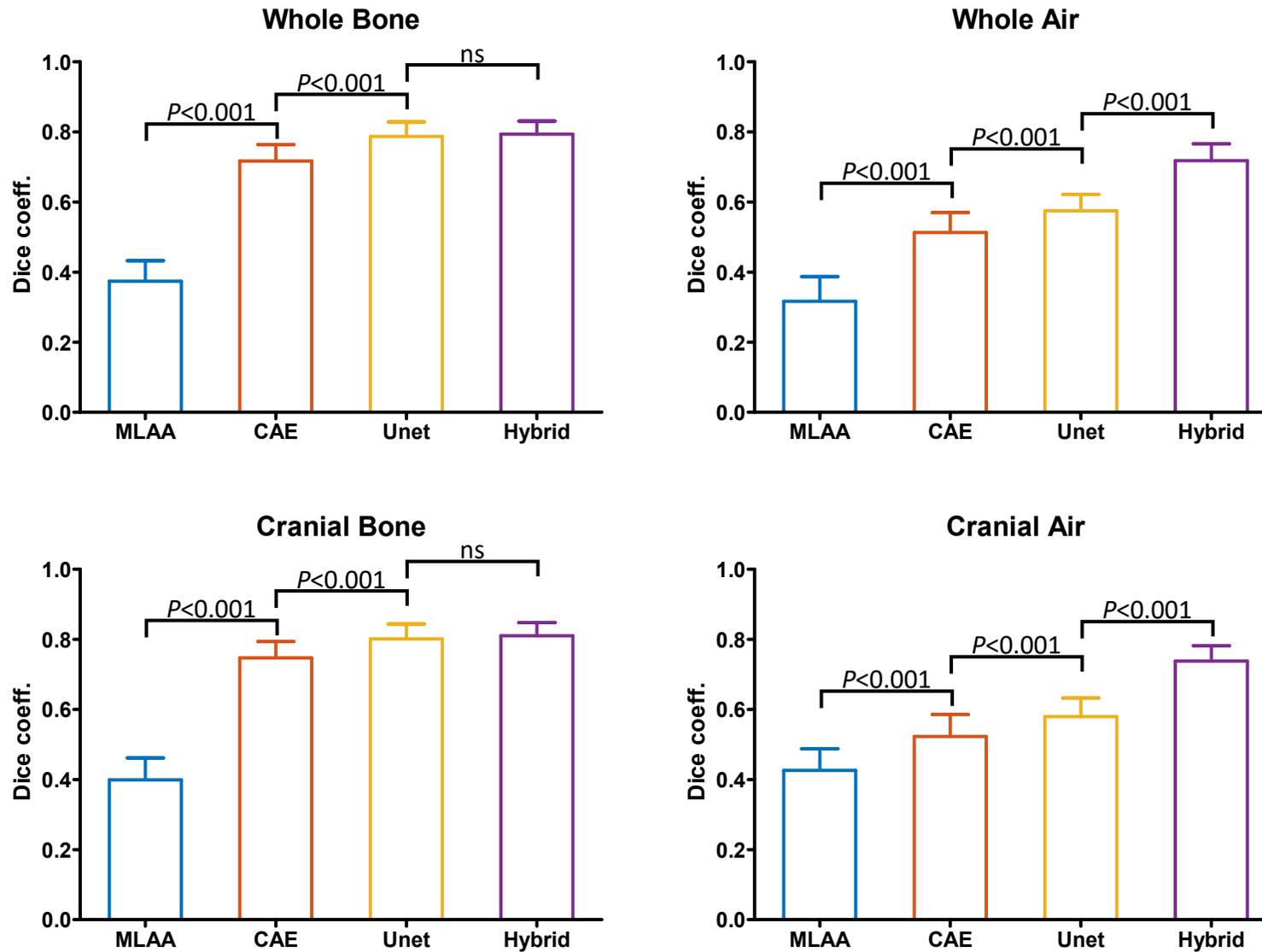
Supplemental Figure 3. Statistical analysis on the Dice similarity coefficients measured in the whole head and only the cranial bone regions for air and bone

Hwang D, Kim KY *et al.*, Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning

μ-Unet μ-hybrid μ-CT
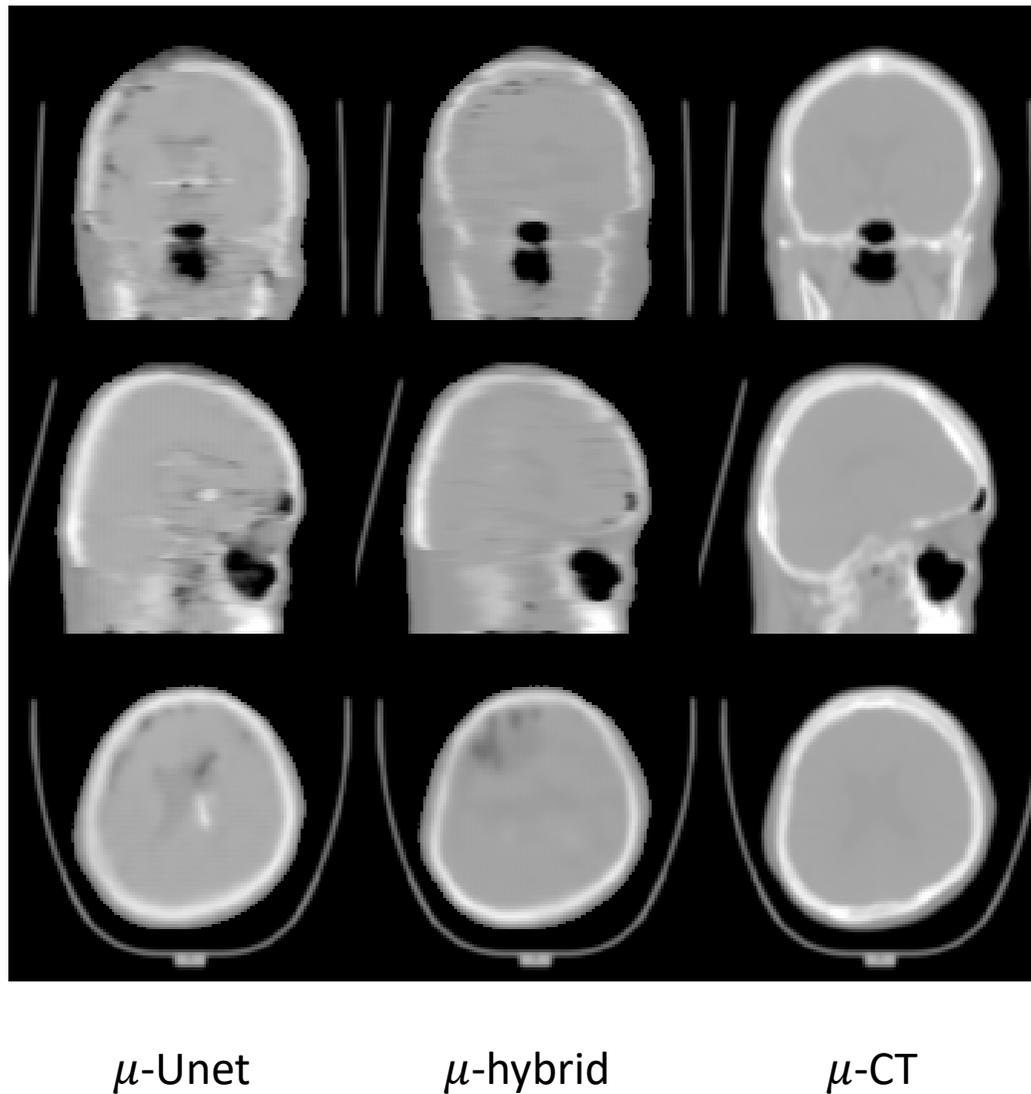
Supplemental Figure 4. CNN (U-net) output (left and middle columns) generated when only the attenuation information is used for CNN training. The right column shows ground truth (attenuation map derived from CT).

Hwang D, Kim KY *et al.*, Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning