



Deep learning for whole-body medical image generation

Joshua Schaefferkoetter^{1,2} · Jianhua Yan³ · Sangkyu Moon² · Rosanna Chan² · Claudia Ortega² · Ur Metser² · Alejandro Berlin^{4,5,6} · Patrick Veit-Haibach²

Received: 20 March 2021 / Accepted: 11 May 2021 / Published online: 22 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Background Artificial intelligence (AI) algorithms based on deep convolutional networks have demonstrated remarkable success for image transformation tasks. State-of-the-art results have been achieved by generative adversarial networks (GANs) and training approaches which do not require paired data. Recently, these techniques have been applied in the medical field for cross-domain image translation.

Purpose This study investigated deep learning transformation in medical imaging. It was motivated to identify generalizable methods which would satisfy the simultaneous requirements of quality and anatomical accuracy across the entire human body. Specifically, whole-body MR patient data acquired on a PET/MR system were used to generate synthetic CT image volumes. The capacity of these synthetic CT data for use in PET attenuation correction (AC) was evaluated and compared to current MR-based attenuation correction (MR-AC) methods, which typically use multiphase Dixon sequences to segment various tissue types.

Materials and methods This work aimed to investigate the technical performance of a GAN system for general MR-to-CT volumetric transformation and to evaluate the performance of the generated images for PET AC. A dataset comprising matched, same-day PET/MR and PET/CT patient scans was used for validation.

Results A combination of training techniques was used to produce synthetic images which were of high-quality and anatomically accurate. Higher correlation was found between the values of mu maps calculated directly from CT data and those derived from the synthetic CT images than those from the default segmented Dixon approach. Over the entire body, the total amounts of reconstructed PET activities were similar between the two MR-AC methods, but the synthetic CT method yielded higher accuracy for quantifying the tracer uptake in specific regions.

Conclusion The findings reported here demonstrate the feasibility of this technique and its potential to improve certain aspects of attenuation correction for PET/MR systems. Moreover, this work may have larger implications for establishing generalized methods for inter-modality, whole-body transformation in medical imaging. Unsupervised deep learning techniques can produce high-quality synthetic images, but additional constraints may be needed to maintain medical integrity in the generated data.

Keywords Artificial intelligence · Deep learning · Attenuation correction · PET · PET/MR

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence).

✉ Joshua Schaefferkoetter
joshua.schaefferkoetter@siemens.com

¹ Siemens Medical Solutions USA, Inc., 810 Innovation Drive
Knoxville TN 37932 USA

² Joint Department of Medical Imaging, Princess Margaret Cancer Centre, Mount Sinai Hospital and Women's College Hospital, University of Toronto, University Health Network, 610 University Ave Toronto Ontario M5G 2M9 Canada

³ Shanghai Key Laboratory for Molecular Imaging, Shanghai University of Medicine and Health Sciences, Shanghai 201318 China

⁴ Radiation Medicine Program, Princess Margaret Cancer Centre, University Health Network, Toronto ON Canada

⁵ Department of Radiation Oncology, University of Toronto, Toronto ON Canada

⁶ Techna Institute, University Health Network, Toronto ON Canada

Background

Recently, artificial intelligence (AI) algorithms based on deep convolutional networks have demonstrated remarkable success for cross-domain image translation, with some of the most impressive results having been produced by systems comprising generative adversarial networks (GANs). Initial work in this field involved natural photographic images, but applications specific to medical imaging emerged soon thereafter [1–3].

This study investigated deep learning transformation for whole-body, medical imaging — demonstrated here for PET/MR applications. Specifically, MR data were used to generate synthetic CT image volumes, which were then used for PET attenuation correction (AC). This approach offers potential advantages over the current default AC methods, which typically use multiphase Dixon sequences to segment various tissue types. Although many improvements have been seen over the last years, Dixon segmentation-based AC is still prone to a variety of errors including inaccurate attenuation values, tissue misclassification and incomplete or misregistered bone atlases.

The idea to use deep learning to improve PET AC has been investigated previously by several different groups with notable progress. One early study used a small population of subjects to train a network to translate MR-to-CT images in a supervised fashion using various loss objectives [4]. This work reported promising results and was specific to the pelvis. Another group used a similar training approach to successfully estimate maps of the attenuating mu values (mu maps) directly from the non-attenuation corrected (NAC) PET images [5] — this approach is interesting because it is not affected by misregistration errors between the PET and an accompanying anatomy image. Another recent work [6] also employed a supervised training approach with paired training data for improving the 3D attenuation maps produced by a maximum likelihood reconstruction of attenuation and activity (MLAA) algorithm [7]. Unsupervised training with unpaired data within a cycle-consistent GAN framework (CycleGAN) [8] has also been investigated for medical imaging [3]. One such study investigated this technique for transforming MR into CT images of the head [9], resulting in high-resolution synthetic sagittal image slices. CycleGAN has also been used in transformations for the whole body [10] — this work incorporated a novel correlation loss to address the issues associated with subject positioning differences between MR and CT. The authors show improvements, but their results were not anatomically accurate across the entire body. The cycle-constrained framework has also been used to translate directly between NAC and AC PET images, mitigating the need for deriving the patient mu map altogether [11].

These prior studies offer important contributions for improving PET AC, but none is without their limitations. Each of these focused on limited anatomical ranges, required sophisticated preprocessing algorithms or produced suboptimal

results which could limit a clinical adoption of the techniques. Algorithms trained on only PET data are prone to anatomical discrepancies and may only be applicable to specific PET tracers. Furthermore, networks trained by supervised, pixel-averaged loss functions, are known to produce relatively blurry outputs, and many were built on networks with 2D architectures not optimized for 3D data.

The experiment detailed here was designed in pursuit of a robust solution for accurate anatomical transformations in whole-body PET/MR AC protocols. This study aimed to investigate the capacity of a GAN system for general MR-to-CT image transformation and to evaluate the quantitative performance of the AI-synthesized images for PET AC. The findings presented here demonstrate the feasibility of this technique and its potential to generate high-quality results which could improve certain aspects of AC for whole-body PET/MR examinations. Moreover, this work may lend its methods to other medical applications in which inter-modality transformations would be helpful.

Methods

Network architecture and training

The deep convolutional networks were trained within a GAN framework, and the performances of two-dimensional and three-dimensional networks were evaluated. The generator and discriminator architectures followed those described in a previous work [12]. The generator comprised sequential residual blocks, situated between encoding and decoding layers at the bottom and the top of the network, respectively. The discriminator followed the patchGAN architecture [13]. The GAN system was trained with adversarial, supervised and unsupervised losses. The supervised objective included a pixel-wise L1-norm (mean absolute error) loss imposed at the output of the generator network [14]. The unsupervised objective included cycle consistency and identity loss terms [8]. This approach required 2 unique generator networks, one for transforming MR to CT and one for CT to MR, designed for mutual regularization. There were also 2 respective discriminator networks, for classifying the real and generated data in each domain, trained with an L2-norm (mean squared error) loss. Both generators and both discriminators were trained in the same way for 15,000 epochs, and used the Adam optimizer. Cross-validation was performed to concurrently monitor training convergence within a separate population of test subjects.

Supervised training objectives require classification information which directly labels the training data — in this case, it meant that a number of paired MR and CT volumes needed to be spatially co-registered. Differences in patient positioning between the 2 scanners made whole-body, global co-registration

challenging, and even impossible. However, local co-registration was used to generate labels at different regions independently. Sub-volumes at various anatomical sites were co-registered and extracted from the patients in the training population. This approach was well-suited for creating training data, since the network was trained with 3D patch samples which were already much smaller than the whole-body volumes.

The supervised and unsupervised training used different datasets — co-registered volume patches were necessary for the supervised objectives, but the unmatched, whole-body volumes were able to be used for the unsupervised training iterations. Although the paired data could be used for both the supervised and the unsupervised training, the unpaired data could not, and it was decided to keep the datasets separate. Hence, the different training approaches were not performed simultaneously, *per se*, but were alternated at each epoch. At every run, the input subject data were randomly augmented by translation, rotation and anisotropic scaling, before randomly extracting a single $96 \times 96 \times 96$ cubic patch from each. Theoretically, this approach yielded an infinite number of unique patch samples available for training — a complete epoch comprised training on 128 samples, with minibatch size 2. For computational efficiency, a separate script to prepare the training data at each epoch ran concurrently on the CPU with the GAN training performed on the GPU.

Training patient population

The underlying transformation task of this work sought to define the mapping specifically between Hounsfield-valued CT and MR Dixon water image domains within the human body. The datasets from 60 patients, imaged with ^{18}F -DCFPyL for evaluation of prostate cancer, were selected for this — every subject gave informed consent for their anonymized data to be used as a part of an institutional REB-approved research study. Each patient underwent separate PET/CT (Discovery MI DR, GE Healthcare) and PET/MR (Biograph mMR, Siemens Healthcare) examinations on the same day. For PET/CT, the CT data were acquired with 120 kVp tube voltage and average current of 165 ± 14.5 mAs; the pixel size of the reconstructed image volumes was 1.3672 mm with slice thickness 3.27 mm. For PET/MR, the MR Dixon data were acquired with Siemens' CAIPIRINHA parallel imaging technique [15]. This sequence is fast and yields high-quality Dixon images with pixel size 1.3021 mm and slice thickness 2.9928 mm.

Whole-body image generation

Once the training was complete, the CT generator network was used to create pseudo CT volumes to be used for PET/MR AC in a set of 30 validation patients. For each subject, the composed whole-body Dixon water image volume was divided into

overlapping patches, which were then processed by the network to produce the corresponding synthesized CT patches. These outputs were then recombined to produce the whole-body volume — an example of this is shown in Fig. 1.

The synthesized CT volumes were converted to 511 keV attenuation mu maps according to the bilinear transformation described in [16].

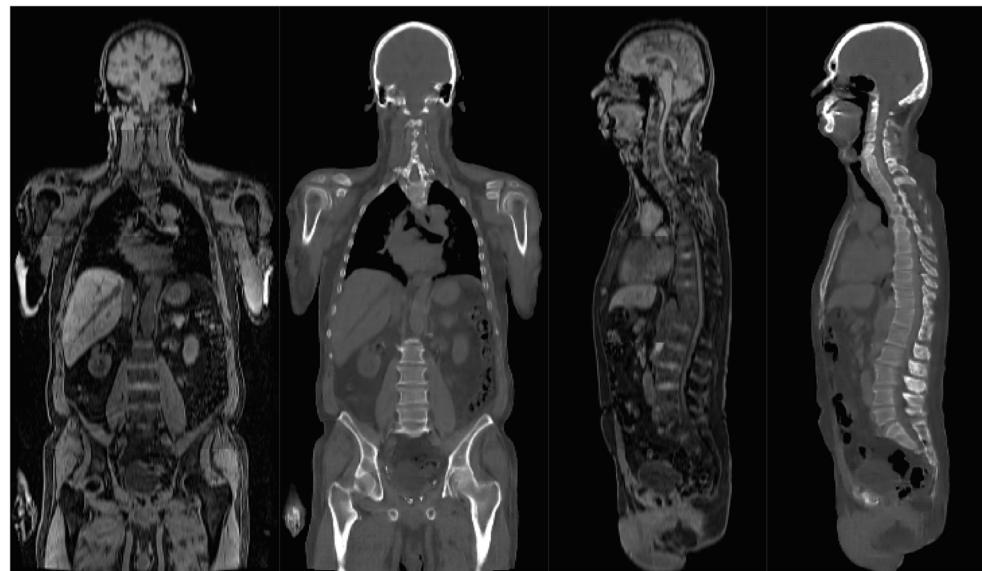
The validation patients received injections of 326.3 ± 14.8 MBq ^{18}F -DCFPyL and were scanned on the PET/MR at 122 ± 7 min post-injection and then on PET/CT at 200 ± 10 min post-injection. In this work, the data from PET/CT were used as the ground truth. As an initial test, the total amounts of attenuating medium contained in both MR-based attenuation maps were compared to those from the CT.

PET evaluation

The PET images reconstructed using different AC mu maps, from the PET/MR, were compared to each other and also to those from the PET/CT. For every patient, the PET/MR data was reconstructed 2 times, once with the default mu map and again with the synthesized CT mu map (synCT) — both were compared to that from PET/CT. In order to account for MR truncation artefacts, MLAA is routinely used at our institution in order to improve PET quantification for all patient scans — we maintained this convention for this work. The MLAA algorithm takes 2 inputs, the incomplete umap and NAC PET data, and from these, simultaneously estimates the most likely distribution of each. The end results here were umaps with “filled in” arms (illustrated for each of the MR-based mu maps in Fig. 6), which were then used for PET AC in the reconstruction. All PET analyses were performed using standardized uptake value (SUV) images to correct for the tracer decay at the different acquisition times. For the reconstructions, the transaxial image pixel dimensions were matched at 2.6 mm, but the slice thickness, which depends on the gantry detector configuration, was 2.03 mm for PET/MR and 3.27 mm for PET/CT. None of the PET reconstructions used time-of-flight information.

Quantitative evaluations were performed for volumes of interest (VOIs) defined at various anatomical locations: liver, lungs, salivary glands and small metastatic lesions. These regions were selected in order to represent a range of tracer uptake characteristics. For the liver and lung, the VOIs were defined by all voxels contained within spheres of 30-mm diameter, manually placed within the organ parenchyma. For the salivary glands, the VOIs were defined by the intracranial voxels having values greater than or equal to 50% of the max. Lesion VOIs were also defined by a 50% max threshold, but since the lesions were much smaller than the salivary glands, the VOIs used smaller spheres drawn over the focal uptake. The VOIs were defined separately on the PET/MR and PET/CT volumes, and the threshold-based voxel selections (for the salivary glands and lesions) were calculated independently for

Fig. 1 A synthetic, whole-body CT volume generated from patient Dixon MR data. Here, the MR and corresponding synthesized CT data are displayed as single slices in coronal (left) and sagittal (right) views



every image. The VOIs are illustrated for 3 representative patients in Fig. 2.

The VOI measurements in the images of both PET/MR AC methods were compared to those of PET/CT. The relative differences in each VOI set were found to be normally distributed by the Shapiro-Wilk normality test. As such, 2-tailed, paired t-tests were used to quantify the significance of any discrepancies between the 2 methods.

Results

Network performance

The convolutional networks were initially trained using only the unsupervised CycleGAN approach, i.e. using only the adversarial,

cycle consistency and identity losses with unpaired training samples. The networks successfully learned the features of each class and produced high-quality, realistic transformations for certain body parts like the head. However, it was observed that these transformations were not anatomically accurate for every region within the whole body — the ribs were incorrectly characterized by the translation, as seen in Fig. 3. Although, this may not have significantly impacted the PET AC in the thorax, we sought to achieve transformations which were anatomically accurate.

Incorporating the supervised loss, with labelled data, into the CycleGAN training resolved this — the results presented throughout this work were produced by only 3D networks trained through this combination. As a sanity check, the performance of the CT generator network, trained using only adversarial and supervised losses, i.e. without unsupervised losses, was visually evaluated, as was that of the corresponding 2D network.

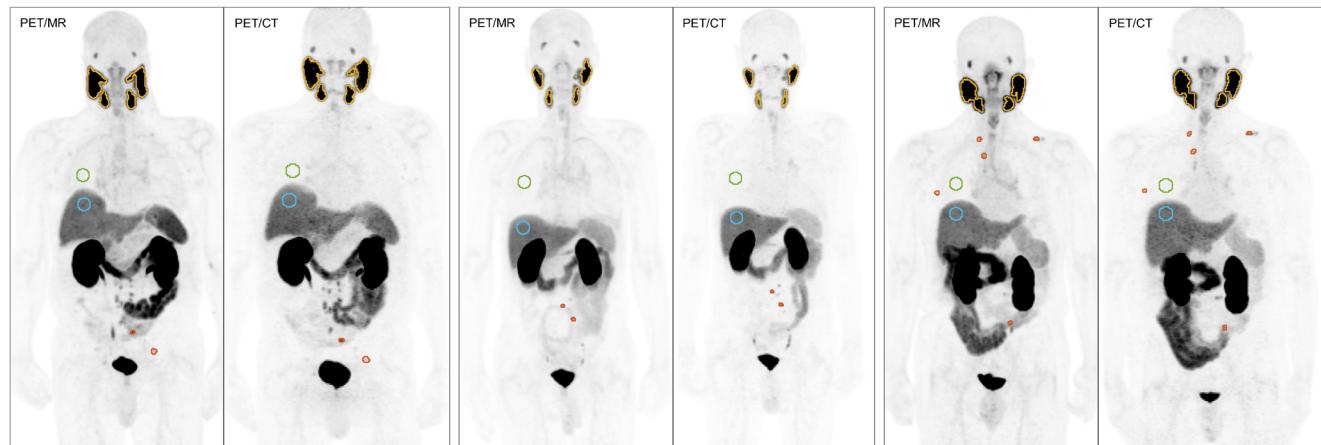
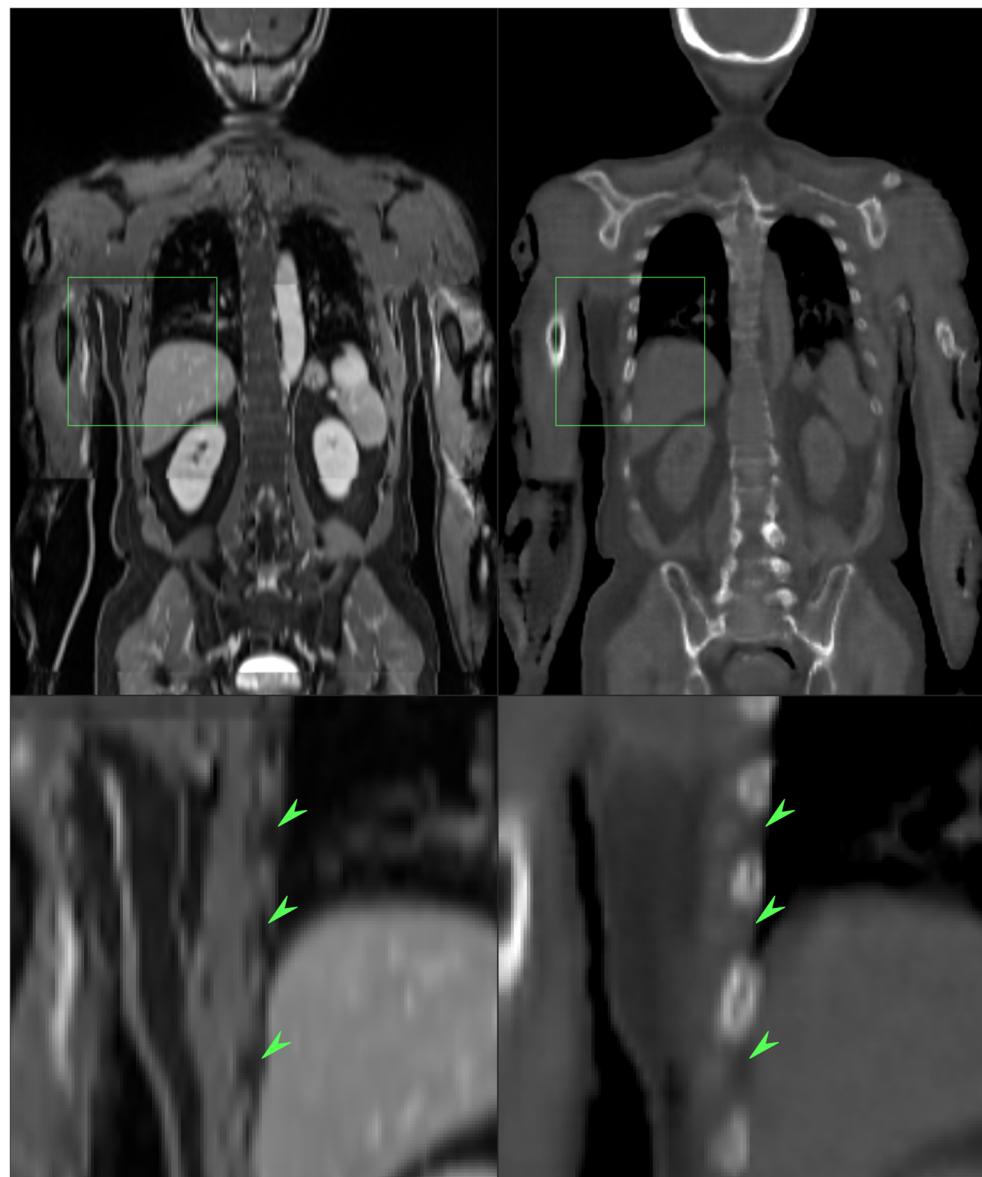


Fig. 2 VOIs were defined at various regions for PET/MR and PET/CT, illustrated here for 3 patients. Liver and lung VOIs, shown in blue and green, were defined by all voxels contained in spheres of 30 mm diameter. The salivary gland VOIs, shown in gold, were defined by the

intracranial voxels, with values greater than or equal to 50% of the max. Lesion VOIs, shown in red, were also defined by a 50% max threshold, but within smaller spheres drawn only over the focal lesion uptake

Fig. 3 A potential pitfall of CycleGAN training. This is an example of inference by a network during training with only unsupervised loss objectives, with the Dixon MR image shown on the left and its corresponding synthesized CT on the right. While this network was successfully learning to reproduce the features of the CT domain and the synthesized images seemedly appeared reasonable, the results were anatomically inaccurate. The zoomed-in views of the outlined regions show the real locations of the ribs (denoted by the arrowheads), which were incorrectly characterized by the transformation. Including supervised losses (with labelled data) during training corrected this



As seen in Fig. 4, all networks were able to learn MR-to-CT translations, but the 2D network yielded whole-body volumes of relatively low overall quality with poor axial contiguity across most of the body. The additional dimension allowed an equivalent 3D network to produce volumes with higher fidelity across all spatial dimensions. Both of these networks were trained using supervised and GAN losses and paired data. Including additional unsupervised objectives with unpaired data introduced substantial improvements for regions which did not have accurate supervised labels, like the hands.

Mu map evaluation

Several advantages were found for the mu maps derived from the synthetic CTs — most notably, the bone maps throughout the entire body were complete with better

anatomical alignment relative to those in the default mu maps. Direct comparison of these whole-body mu maps with those of the CT was challenging due to patient positioning and the resulting complex misregistration. However, in the head, where simple co-registration was possible, a higher correlation of the quantified mu values was observed for the synCT mu maps — Fig. 5 illustrates this. The top two rows show identical line profiles drawn over the default and synCT umaps resulted in mean squared errors of 0.35 and 0.15, respectively, relative to the CT umap. This slice was chosen to also highlight a characteristic pitfall often encountered in the default mu maps, that is, incorrectly assigning tissue values to air within the intracranial sinuses. The correlations between all voxels within the head are shown on the bottom row of the same figure, along with the linear regression fits.

Fig. 4 A visual comparison of the transformations produced by a 2D network (on the left) and that by an equivalent 3D network (in the middle) — the additional spatial dimension of the 3D network substantially improved the quality of the inference. Both of these networks were trained using the same supervised L1-norm and GAN adversarial losses. The transformation on the right was produced by a 3D network trained with additional unsupervised adversarial, cycle consistency and identity losses. These additional objectives improved the quality further; this is especially notable in regions like the hands, where difficult co-registration prevented accurate supervised labels



Indeed, the synCT resulted in much higher Pearson correlation coefficient between CT mu values ($PCC = 0.885$), with higher coefficient of determination (slope = 0.8; $R^2 = 0.78$), relative to that of the default mu map ($PCC = 0.651$; slope = 0.55; $R^2 = 0.42$) for all voxels included within a mask of the entire head. The significance values displayed on the scatter plots correspond to the F statistic of the linear regression.

PET evaluation

The default and synCT attenuation maps from the PET/MR were then used for AC in PET reconstructions. The two resulting PET images were compared directly to those from the PET/CT in the evaluation subjects. An overview of this is presented in Fig. 6.

For each subject, the axial fields of view were matched, and the total amounts of attenuation and tracer activity throughout the body were measured. The biases are shown for each MR-based map in Fig. 7. Both MR-derived mu maps underestimated the total amount of attenuation, but the additional bony regions in the mu map derived from the synthesized CT reduced this negative bias. As a result, the total amounts of reconstructed PET activities were slightly greater with the synCT mu maps. In both cases, the quantitative differences between MR-AC methods were found to be significant at the 5% level.

The measurements of tracer activities were performed on PET images taken at different scanning points and uptake times (~122 min P.I. for PET/MR and ~200 min P.I. for PET/CT) — as demonstrated in Fig. 7, this had little effect

on the total amount of measured tracer in the body. The top row of Fig. 8 shows the absolute differences between local SUV measurements in the PET/MR and PET/CT images. These were expected to be similar, and it is seen here that the median absolute differences in the VOI measurements were lower in the images reconstructed with the synCT mu maps relative to those reconstructed with the default mu maps, though the differences between methods were not significant except for in the case of the salivary glands. These data are presented in the figures as percentage differences relative to PET/CT as ground truth.

Accurate interpretation of the total differences in regional tissue measurements, however, is somewhat more complicated — different tissues will have different tracer uptake and washout properties. Both mu map methods produced images which generally followed the expected trends, with exception of the salivary glands. In this region, the synCT mu maps produced, presumably, more quantitatively accurate images with systematically lower measurement difference. In fact, it was only in this region in which the differences between MR-AC methods were statistically significant.

Discussion

This study investigated the potential of 3D deep convolutional networks for cross-domain, medical image translation. In particular, it focused on whole-body transformation, and in this context, state-of-the-art results were achieved.

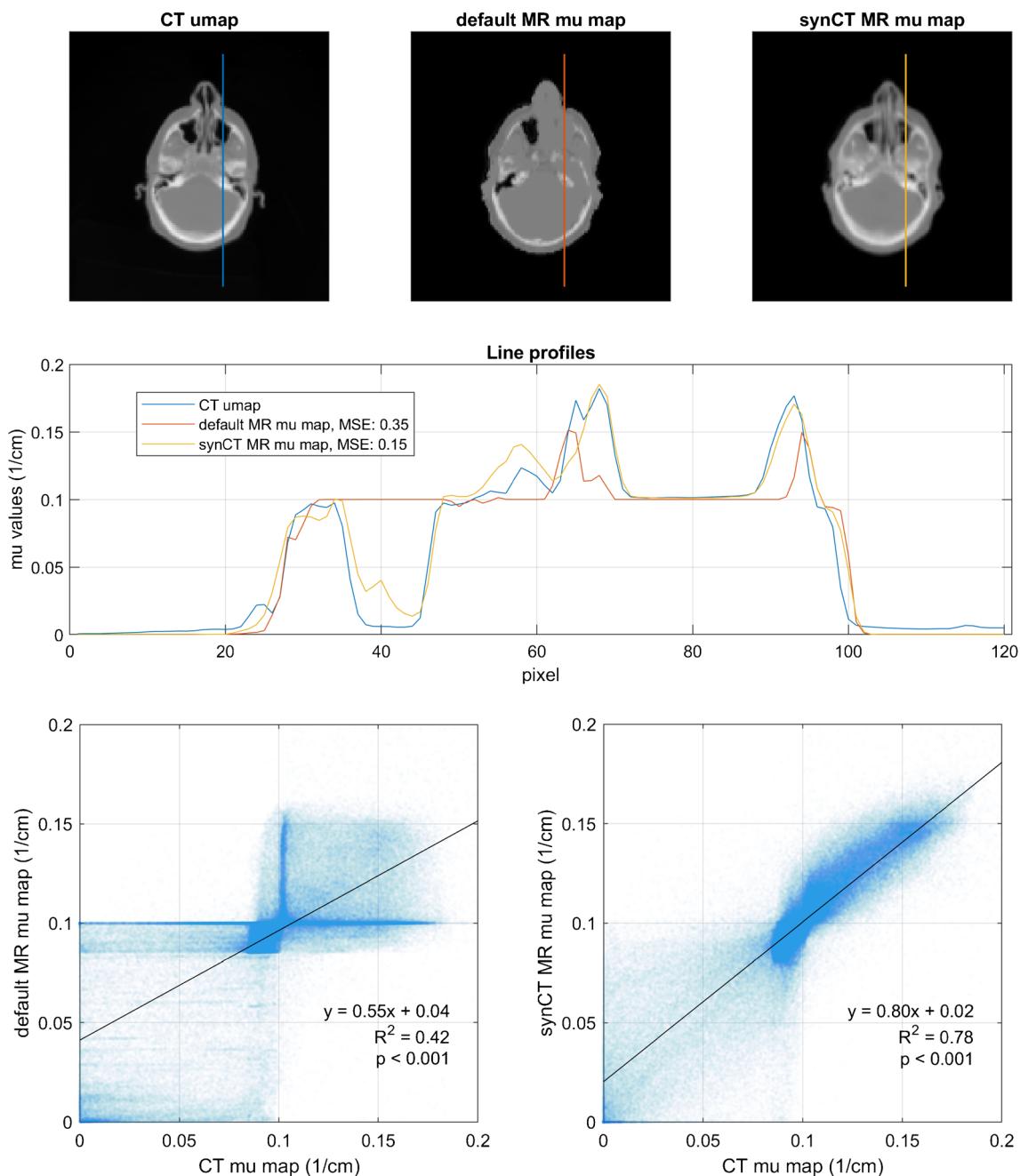


Fig. 5 A representative example showing the head mu maps of single subject, co-registered between PET/CT and PET/MR. The top two rows show the profiles of identical lines drawn over three mu maps. Taking the CT-derived mu map as the ground truth, the mean squared error was less among mu values in the synCT mu map than those in the default mu map.

The bottom row shows the correlations for all voxels within the masked whole head of the same patient, along with the linear regression fits. The mu values within the synCT mu map showed a significantly higher correlation with the CT mu values than did those of the default mu map

The novelty of this work lies in several aspects. It has been previously shown that sophisticated deep learning systems trained on unpaired data are capable of producing high-quality synthetic images, but potential pitfalls of using such an approach for medical applications are less documented. This study found that additional constraints were needed to generate structurally accurate data. The GAN system here was trained using both paired and unpaired training data, allowing

a unique combination of supervised and unsupervised loss objectives. This combined approach yielded high-quality synthetic CT data which were found to be anatomically correct. The convolutional networks used here were built on 3D architectures. This improved the translational quality of the volumetric data over 2D networks. A unique set of whole-body patient data was used to evaluate the networks' performance for improving PET attenuation correction, and image

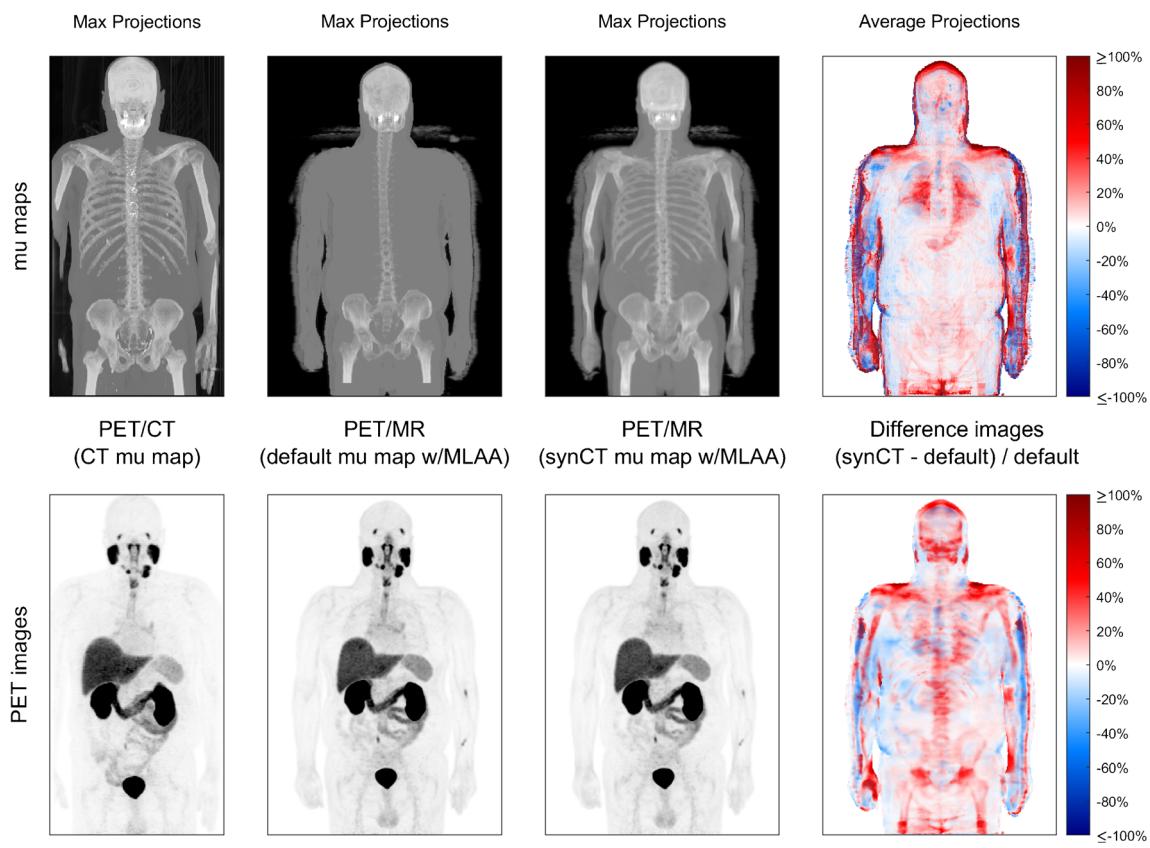


Fig. 6 Whole-body PET reconstructions on the PET/MR using the default (2nd column) and synCT (3rd column) mu maps were performed in a set of validation patients and compared to those from PET/CT (1st column). In this figure, the different AC mu maps are shown on the top

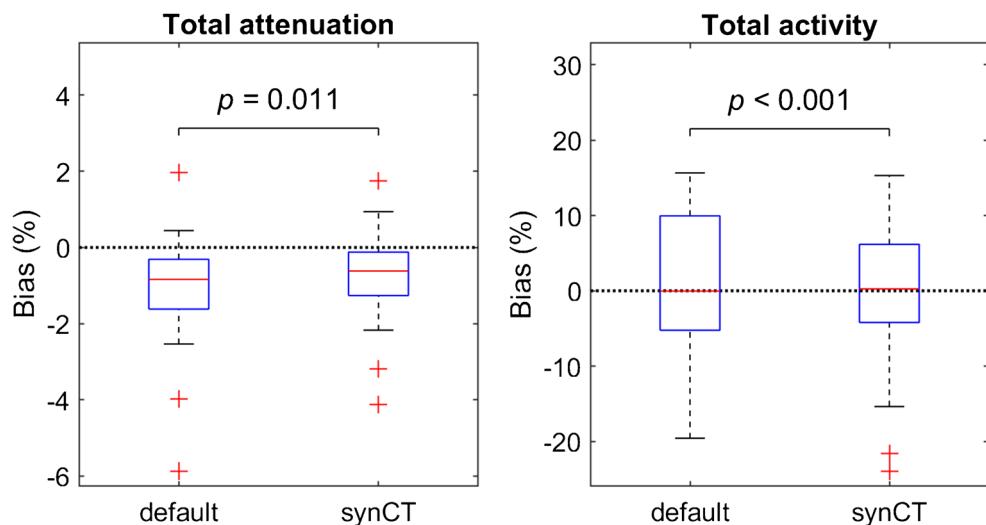
row and the resulting PET reconstructions on the bottom. The last column shows the relative differences within the body between the 2 MR-AC approaches

quantification was compared to a set of matched, same-day PET/CT reconstructions.

The main goal of this work was to demonstrate the efficacy of this whole-body 3D approach for general whole-body transformation tasks. The results are promising but must be interpreted cautiously — it would be wrong to assert that any

AI-synthesized image has inherent clinical value on its own. For example, it might not be possible to produce a T2-weighted image, generated from a T1-weighted image, which could be used for accurate pathological diagnosis, i.e. the T1-weighted image may not provide sufficient information to inform accurate mapping to the T2 domain.

Fig. 7 PET/MR measurement differences relative to PET/CT for the total attenuation in the mu maps and corresponding reconstructed PET activities. The data in each plot are presented as biases, since, ideally, they are expected to be zero. The mean population differences between both methods were significant at the 5% level



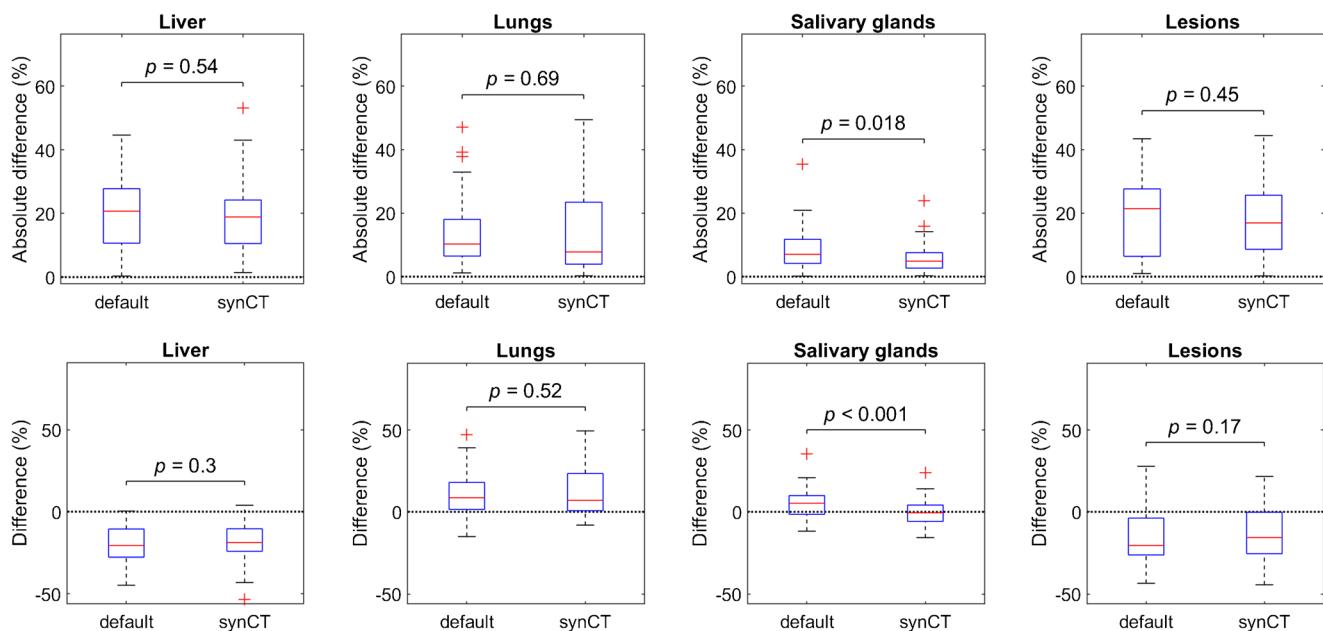


Fig. 8 Absolute and total PET measurement differences for both MR-based mu maps, relative to PET/CT, for VOIs located in different anatomical regions (illustrated in Fig. 2). The top row shows that the median absolute differences were consistently lower when using the synCT mu maps. None of these or the total differences shown in the bottom row was expected to be exactly zero, however, due to the different scanning time points between

PET/MR and the later PET/CT. Instead, it was expected that differences in the regions which express PSMA, i.e. liver, glands and lesions should not be positive. In contrast, measurement of the lung tissue, which does not express PSMA, comprises mainly blood pool activity and should not show negative differences. The paired t-test analyses indicated that the performances of both methods were not significantly different, except for in the salivary glands

Such AI transformation techniques are immediately more useful in situations in which the real data provide the complete set of information needed for the inference. In the current experiment, the whole-body MR volume provided the anatomical template from which characteristic bone structures were generated. In other words, although the synthesized CT data is not likely sufficient to diagnose bone disease, we found that they do provide a comprehensive and realistic map of Hounsfield values. Overall, AC for PET/MR systems seems like a well-suited application, and the performance evaluation of the synthesized data presented here was investigated within this context. The quantification within the reconstructed PET images was compared to that within the images processed using the conventional MR-AC method, using PET/CT as the reference. This analysis required certain considerations regarding tracer uptake characteristics in various regions, due to the different scanning time points between PET/MR and PET/CT. The results suggest possible areas of improvement using the new method.

The PET AC evaluation showed that the two methods for estimating attenuation maps performed similarly in some regards. Although the total amounts of attenuation were more accurately estimated with the AI-generated mu map, the median total amounts of reconstructed whole-body PET activities were not substantially different between the two methods. The latter point, of course, depends on the distribution of PET tracer used, which in this case was the prostate-specific membrane antigen (PSMA) agent ^{18}F -DCFPyL. If, instead, a tracer was used with a larger distribution adjacent to the bones, e.g.

^{18}F -NaF, we would expect a larger difference between the total amounts of corrected PET activities.

Notwithstanding this, analyses of the regional measurements revealed some differences. Since the majority of the tracer uptake, due to irreversible tracer-receptor binding, should have already occurred in the 2 h before the first scan [17], tracer activity concentrations in every tissue would be expected to be roughly similar between both scanning time points. The top row of Fig. 8 shows the absolute differences between local SUV measurements in the PET/MR and PET/CT images, and the median absolute differences in the VOI measurements are lower in the images reconstructed with the synCT mu maps relative to those reconstructed with the default mu maps. The total differences for every region (seen in the bottom row) must be interpreted while considering the physiological PSMA expression in each tissue. Tracer activity concentrations in tissues known to express PSMA, i.e. liver, glands and metastatic lesions, were not expected to decrease in the 2nd scan. In contrast, measurements in the lung tissue mainly comprise unbound, circulating tracer in the blood and therefore should not increase in the 2nd scan. The results showed that both MR-AC methods produced images which generally satisfied these expectations for the measured tissues, with the exception of the salivary glands. In this region, the AI-generated mu maps produced PET images with consistently lower SUV measurements.

In this study, obvious differences were observed through direct comparisons of the mu maps, and in this regard, clear advantages were realized by the synCT mu maps. It was challenging,

however, to identify a viable approach for PET validation, especially since the “ground truth” data were from a PET/CT scanner with different acquisition characteristics. This subjected the analyses to potential bias, since different gantry design and processing techniques can lead to inconsistencies in reconstructed activity measurements, even regardless of AC. Considering this, efforts were made to ensure that the reconstructions were similar, e.g. both incorporated system resolution modelling with similar numbers of iterative updates, the transverse pixel sizes were matched and identical smoothing kernels were used. Notwithstanding this, inconsistencies were still inevitable. Hence, the findings were presented here as generalized trends of the subject population, under the assumption that both scanners were calibrated and quantitatively accurate. This assumption is reasonable since both scanners were used clinically and underwent frequent quality control.

The findings of this study revealed that the AI-based AC method might offer potential improvements for local PET quantification in certain anatomical regions — this was observed here for the salivary glands, which seemed to be over-corrected by the conventional method. However, the PET quantification in other regions could also be improved by this method. For example, focal bony metastases would likely realize significant benefit from more complete and accurate bone information in the mu map. The patient population used in this work was scanned for primary staging and did not contain a large number of these lesions, but this would be an interesting direction for future work.

Conclusion

This study demonstrated the possibility for leveraging AI techniques to improve certain aspects of MR-based PET attenuation correction. We demonstrated that whole-body 3D MR image volumes can be transformed into synthetic CT image volumes for use in PET AC with high accuracy. However, this work may have larger implications for inter-modality, medical image transformation tasks in general. Similar methods could be applied to other aspects of whole-body imaging, potentially opening the door to a new set of AI-based clinical applications.

Declarations

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Conflict of interest The first author JS is a full-time employee of Siemens Medical Solutions, USA. The other authors declare no competing interests.

References

- Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal.* 2019;58:101552.
- Dar SU, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging.* 2019;38(10):2375–88.
- Armanios K, et al. Unsupervised medical image translation using Cycle-MedGAN. In 2019 27th European Signal Processing Conference (EUSIPCO). 2019. IEEE.
- Leynes AP, et al. Zero-echo-time and Dixon deep pseudo-CT (ZeDD CT): direct generation of pseudo-CT images for pelvic PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI. *J Nucl Med.* 2018;59(5):852–8.
- Armanios K, et al. Independent attenuation correction of whole body [¹⁸ F] FDG-PET using a deep learning approach with Generative Adversarial Networks. *EJNMMI Res.* 2020;10:1–9.
- Hwang D, et al. Generation of PET attenuation map for whole-body time-of-flight ¹⁸F-FDG PET/MRI using a deep neural network trained with simultaneously reconstructed activity and attenuation maps. *J Nucl Med.* 2019;60(8):1183–9.
- Michel, C.J. and J. Nuyts, Completion of truncated attenuation maps using maximum likelihood estimation of attenuation and activity (MLAA). 2013, Google Patents.
- Zhu, J.-Y., et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. in Proceedings of the IEEE international conference on computer vision. 2017.
- Wolterink, J.M., et al. Deep MR to CT synthesis using unpaired data. In International workshop on simulation and synthesis in medical imaging. 2017. Springer.
- Ge, Y., et al. Unpaired whole-body MR to CT synthesis with correlation coefficient constrained adversarial learning. In Medical Imaging 2019: Image Processing. 2019. International Society for Optics and Photonics.
- Dong X, et al. Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. *Phys Med Biol.* 2020;65(5):055011.
- Johnson, J., A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision. 2016. Springer.
- Li, C. and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In European conference on computer vision. 2016. Springer.
- Isola, P., et al. Image-to-image translation with conditional adversarial networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- Breuer FA, et al. Controlled aliasing in parallel imaging results in higher acceleration (CAIPRINHA) for multi-slice imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine.* 2005;53(3):684–91.
- Carney JP, et al. Method for transforming CT images for attenuation correction in PET/CT imaging. *Med Phys.* 2006;33(4):976–83.
- Szabo Z, et al. Initial evaluation of [¹⁸F] DCFPyL for prostate-specific membrane antigen (PSMA)-targeted PET imaging of prostate cancer. *Mol Imaging Biol.* 2015;17(4):565–74.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.