medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Medical Image Analysis (2022)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Enhanced Direct Joint Attenuation and Scatter Correction of Whole-Body PET Images via Context-Aware Deep Networks

Saeed Izadi[a], Isaac Shiri[b], Carlos F. Uribe[c,d], Parham Geramifar[e], Habib Zaidi[b,g,h,i], Arman Rahmim[d,f,j], Ghassan Hamarneh[a]

[a]*Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Canada*
[b]*Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva 4, Geneva, Switzerland*
[c]*Functional Imaging, BC Cancer, Vancouver, BC, Canada*
[d]*Department of Radiology, University of British Columbia, Vancouver, Canada*
[e]*Research Center for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran*
[f]*Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, BC, Canada*
[g]*Geneva University Neurocenter, Geneva University, Geneva, Switzerland*
[h]*Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands*
[i]*Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark*
[j]*Departments of Physics, University of British Columbia, Vancouver, BC, Canada*

ARTICLE INFO

Article history:

Keywords:
PET/CT
Whole-body
Attenuation Correction
Deep learning

ABSTRACT

In positron emission tomography (PET), attenuation and scatter corrections is necessary steps towards accurate quantitative reconstruction of the radiopharmaceutical distribution. Inspired by recent advances in deep learning, many algorithms based on convolutional neural networks have been proposed for automatic attenuation and scatter correction, enabling applications to CT-less or MR-less PET scanners to improve performance in the presence of CT-related artifacts. A known characteristic of PET imaging is to have varying tracer uptakes for various patients and/or anatomical regions. However, existing deep learning-based algorithms utilize a fixed model across different subjects and/or anatomical regions during inference, which could result in spurious outputs. In this work, we present a novel deep learning-based framework for direct reconstruction of attenuation and scatter corrected PET from non-attenuation-corrected images in absence of structural information. To deal with inter-subject and intra-subject uptake variations in PET imaging, we propose a novel model to perform subject- and region-specific filtering through modulating the convolution kernels in accordance to the contextual coherency within the neighboring slices. This way, the context-aware convolution can guide the composition of intermediate features in favor of regressing input-conditioned and/or region-specific tracer uptakes. We also utilize a large cohort of 910 whole-body studies for training and evaluation purposes, which is more than one order of magnitude larger than previous works. In our experimental studies, qualitative assessments showed that our proposed CT-free method is capable of producing corrected PET images that accurately resemble ground truth images corrected with the aid of CT scans. For quantitative assessments, we evaluated our proposed method over 112 held-out subjects and achieved absolute relative error of $14.30 \pm 3.88\%$ and relative error of $-2.11\% \pm 2.73\%$ in whole-body.

© 2022 Elsevier B. V. All rights reserved.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1. Introduction & Related Work

^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging is one of the leading imaging modalities for quantitative *in vivo* measurement of physiological and biochemical processes with applications in oncology (Rohren et al., 2004), cardiology (Dilsizian et al., 2016) and neurology (Herholz and Heiss, 2004). Studies reveal that 40-50% of the recorded annihilation coincidences in a typical whole-body PET scan are affected by Compton scatter where one or both photons deviate from the original line of response before reaching the PET detectors (Zaidi and Koral, 2004). The scattered photons may interact with other surrounding dense material, such as bed instruments, electronic components, and patient body, and carry inaccurate information about the annihilation site (Gong et al., 2019). Also, scattered photons may fall out of the predefined energy windows and fail to be recorded by the PET detectors, which leads to photon attenuation (Gong et al., 2019). In either case, inaccurate number of detected coincidences contributes to under-estimation or over-estimation of the tracer distribution, which ultimately results in inaccurate uptake quantification and notable visual artifacts in the reconstructed PET images. These factors can complicate the assessment of PET findings and may lead to incorrect (false negative/positive) clinical diagnoses. Accordingly, it is crucial to correct for the loss of annihilation photons (attenuation correction) and the inaccurately recorded coincidences (scatter correction) during image reconstruction (Boellaard, 2009).

With the introduction of PET/CT (Kinahan et al., 1998), and more recently PET/MR scanners (Fei et al., 2009), high-resolution anatomical images from MRI and CT images are leveraged to derive attenuation and scatter information to obtain attenuation-and-scatter-corrected PET images (PET-AC) (Gong et al., 2019). In hybrid PET/CT scanners, the Hounsfield units recorded in CT scans can be directly converted into the PET 511-keV linear attenuation coefficients. Despite their popularity in routine clinical practice, the CT-based correction methods have several drawbacks including the risk of artifact propagation and position mismatch between PET and CT scans (Berker and Li, 2016) and the exposure to ionizing radiation; of particular concern for pregnant and pediatric patients (Cheuk et al., 2012).

Alternatively, PET/MR is a non-ionizing imaging technique that allows enhanced visualization of soft tissue contrast without necessitating radiation exposure (Vandenberghe and Marsden, 2015). However, aside from the relatively more limited availability of MR scanning time, the MRI tissue intensities cannot be directly converted into 511 keV linear attenuation coefficients as in CT-based systems (Yang et al., 2019b). To leverage MRI for PET attenuation correction, segmentation and registration-based techniques have been developed. In segmentation-based techniques, uniform linear attenuation coefficients are assigned to different tissues inferred from segmenting an MRI image (Arabi et al., 2015; Berker et al., 2012). In the registration-based methods, an atlas of pre-acquired CT images is used to obtain an attenuation map template, which is then spatially registered to match the patient's body using the anatomical information in the MRI (Arabi et al., 2016). In addition to the computational overhead, these techniques may deteriorate PET reconstruction due to tissue mis-classification, imprecise co-registration, data truncation and metal-induced susceptibility artifacts in CT and/or MR (Lillington et al., 2020; Gong et al., 2019).

Furthermore, enabling CT/MRI-less PET scanners can open up new possibilities for more compact, affordable PET-only scanners of the future (e.g. used for screening, etc.). To avoid relying on structural information captured by CT or MRI, an attempts was made to simultaneously reconstruct the tracer distribution and attenuation maps from the emission data only using maximum likelihood estimation of uptake and attenuation (MLAA) (Nuyts et al., 1999). However, MLAA suffers from slow convergence and low signal-to-noise ratio (SNR) (Mehranian and Zaidi, 2015). Advances in electronics and scintillation research have augmented some commercial PET scanners with time-of-flight (TOF) information. The use of TOF information expedites the convergence time of MLAA and improves SNR (Defrise et al., 2012). Nevertheless, the shortcomings of existing TOF-based algorithms persist mainly due to the uncertainty in detecting the true event positions given the limited time resolution of TOF and high noise.

In recent years, the popularity of deep learning (DL) has ignited extensive research aimed at leveraging the capabilities of neural networks in medical imaging (Litjens et al., 2017; Yedder et al., 2021; Zaidi and El Naqa, 2021). The representative works for attenuation and scatter correction (ASC) using convolution neural networks (CNN) include generation of pseudo-CT images from MR sequences such as ultra-short echo time (UTE) (Leynes et al., 2018), zero echo time (ZTE) (Gong et al., 2018; Blanc-Durand et al., 2019), and Dixon sequences (Leynes et al., 2018). There have been some recent efforts to directly produce pseudo-CT images from non-attenuation-corrected images (PET-NC) (Dong et al., 2019) or uptake and attenuation maps obtained from MLAA (Hwang et al., 2018, 2019). In addition, several studies have employed more advanced DL-based techniques, such as generative adversarial networks (GAN), to boost the performance of pseudo-CT synthesis (Choi and Lee, 2018; Tao et al., 2020). Specifically, MedGAN framework (Armanious et al., 2020b) was utilized to approximate PET-NC to pseudo-CT mapping for whole-body images using paired training data (Armanious et al., 2020a). Recently, methods based on 3D cycleGAN have been applied to pseudo-CT synthesis with the goal of circumventing the need for paired training sets and improving pseudo-CT images by enforcing inverse consistency (Lei et al., 2020a). Despite their substantial improvements over traditional algorithms, methods based on GAN are prone to hallucinating imprecise features in the images (Cohen et al., 2018) and consequently deteriorate the performance of the downstream ASC task (Seith

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

et al., 2017).

Another family of approaches that have recently thrived in tackling direct ASC is emission-only (PET-NC to PET-AC) techniques. These approaches only operate on the PET images or sinograms without requiring any anatomical information. The merits of the encoder-decoder networks in direct PET-NC to PET-AC mapping was demonstrated for brain (Shiri et al., 2019) and whole-body (Shiri et al., 2020) images. Most recently, Arabi et al. (Arabi and Zaidi, 2020) trained a multi-input network to predict the attenuation coefficient factors for a reference slice from its different TOF sinogram bins pertinent to the same slice. Built upon the cycleGAN framework, some works proposed to improve PET-NC to PET-AC mapping by restricting the search space through introducing a reverse network for PET-AC to PET-NC (Dong et al., 2020; Lei et al., 2020b).

Most of the DL-based algorithms for direct attenuation correction adopt either 2D or 3D networks. In the former, either individual 2D patches or full-resolution slices along the axial dimension are processed by a sequence of 2D convolutional layers while in the latter the entire volumetric input is manipulated via 3D convolutional layers during the training and inference. Even though 2D models generally benefit from relatively larger training data (i.e. slices), they fail to integrate the contextual coherency within neighboring slices. On the other hand, 3D networks provide the added value of accessing cross-slice contextual information at the expense of fewer training samples, higher computational cost, and increased number of learnable parameters that can amplify the chances of over-fitting. Furthermore, the information from distant organs are often needlessly aggregated when volumetric inputs are processed by 3D networks. An attempt has been made to use 2.5D training schemes for low-dose PET denoising where multiple consecutive slices are fed into the network to account for cross-slice contextual information (Xu et al., 2017). However, a potential drawback of 2.5D networks is that the contextual information is aggregated in the very first layer of the network and thus it is less likely for such information to influence deeper layers that have more direct impact on the final reconstruction outcome.

Another prominent characteristic of PET images is that the distribution of the tracer uptake tends to undergo considerable inter-subject variations due to changes in factors such as the administered dose, imaging time, and reconstruction parameters. Even the imaging noise and partial volume effects can also magnify the tracer variations across subjects (Gong et al., 2019). On the other hand, the tracer uptake may also vary substantially from one organ to another, even though it is nearly homogeneous inside individual anatomical regions. Despite such input-dependent variations, the existing DL-based methods adopt traditional convolution layers to solve ASC problem wherein the learned convolution kernels are kept fixed over different subjects and body organs during training and inference.

Recent research in dynamic neural networks (Jia et al., 2016; Su et al., 2019; Lin et al., 2020) and guided filtering (Tang et al., 2020; Wang et al., 2018) are promising attempts to produce context-aware networks. To perform object/activity recognition from natural images and machine translation, Lin et al. (Lin et al., 2020) proposed context-gated convolutions to adapt the convolution kernels based on the statistics of the incoming feature representation. Inspired by these works, we provide an alternative way to incorporate adequate 3D contextual information into 2D networks for CT/MRI-less ASC through introducing context-aware convolutional layers (CAC). Our proposed model differs from what mentioned above by (1) the 3D contextual information within neighboring slices is used to adaptively modulate the kernels in specific layers and (2) the network is still in 2D manner and therefore enjoys accessing to large training instances as well as light computation cost at inference. In particular, the neighboring slices are fed into a shared sub-network to extract the region-wise information, which is subsequently used to modify the convolution kernels in an adaptive manner. We use context-aware convolutions only in the decoder layers instead of the entire network to achieve superior performance and efficiency trade-off. Patch-based training does not respect the physics of the attenuation and scatter artifacts and therefore our proposed network operates on full-resolution 2D slices in the input layer.

We address the limitations of previous PET-only ACS works by making the following contributions.

- In contrast to 3D and 2.5D networks, we propose a 2D network that exploits guidance from neighboring slices to propagate input-conditioned context in the convolution kernels.
- We propose to adopt a shared sub-network to extract context information from the set of neighboring slices, which is then used for convolution modification.
- We leverage channel and spatial interactions to adapt the convolution kernels based on the extracted contextual information.
- We conduct our study on a large cohort of 910 subjects that is more than one order of magnitude larger than previous works (Shi et al., 2019; Yang et al., 2019a; Hwang et al., 2019; Lei et al., 2020b).

2. Materials

2.1. Subjects and PET/CT Acquisition

Whole-body ^{18}F-FDG PET/CT scan data of 910 subjects (demographics shown in Table 1) were acquired between 2016-2018 on a Siemens Biograph 6 True point scanner. To avoid any interventional effect on diagnosis, treatment or management of patients, retrospective use of the scan data and waiver of consent was approved by Institutional Review Board (ethic number IR.TUMS.MEDICINE.REC.1398.525) of our institute

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1
Patient demographics of clinical whole-body PET/CT studies enrolled in this study and train/test split information.

	Sex	No.	Age	Weight (kg)	Injected Dose (MBq/g)	Time Post-Injection
			mean ± std (min, max)	mean ± std (min, max)	mean ± std (min, max)	mean ± std (min, max)
Train	Male	422	48 ± 18 (6, 87)	76.1 ± 16.4 (21.0, 145.0)	374.7 ± 55.1 (123.2, 488.4)	60 ± 14 (0, 238)
	Female	376	49 ± 15 (2, 85)	69.5 ± 15.0 (11.5, 138.0)	369.9 ±45.3 (177.6, 555.0)	59 ± 10 (0, 86)
Test	Male	61	45 ± 17 (11, 74)	75.9 ± 14.4 (25.0, 113.0)	366.4 ± 45.2 (170.2, 469.9)	60 ± 11 (8, 120)
	Female	51	50 ± 13 (23, 74)	71.7 ± 12.0 (50.0, 106.0)	367.0 ±34.7 (284.9, 444.0)	62 ± 17 (2, 160)

(Tehran University of Medical Science). All patients were intravenously administered 367.74 ± 48.87 of ^{18}F-FDG and the PET/CT scanning was performed after an uptake period of 60 ± 13 minutes. To obtain attenuation-corrected ground truth PET images that leverage CT-based attenuation maps (PET-CT), a low-dose CT scan (110 kVp, 145 mAs) was conducted before the PET scanning. PET images were reconstructed using the ordinary Poisson ordered subsets-expectation maximization (OP-OSEM) algorithm with 2 iterations and 21 subsets followed by a 5-mm FWHM Gaussian post-reconstruction smoothing. Point-spread functions (PSF) were incorporated in the reconstruction procedure. The matrix size of the reconstructed images was 168 × 168 with a voxel size of 4.073 × 4.073 × 3 mm^2.

2.2. Data Preparation

From the entire cohort, 798 subjects were randomly selected and split into training and validation cohorts with 95:5 ratio, respectively. The validation set was used to select appropriate values for training hyper-parameters and to monitor the risk of over-fitting during the training. For loss calculation, the dynamic range of the PET-NC and PET-CT images were converted to standardized uptake value (SUV) based on the injected dose, decay factor and patient's weight available from the DICOM headers. In our implementation, we used axial slices and cropped each to 152 × 152 pixels to shrink the background (without cropping any foreground) and to emphasize the foreground tracer uptakes. For clinical evaluations, a held-out test cohort consisting of 112 subjects were used. Table 1 summarizes the patient demographics for training and test cohorts.

3. Methods

We first propose the mathematical formulation of direct ASC as a supervised image translation problem and introduce the necessary notations. For the solution, i.e. obtaining PET-AC outputs from PET-NC inputs, we adopt a U-Net (Ronneberger et al., 2015) architecture and equip it with the context-aware convolutional layers (CAC). The network is designed to deliver a solution with specific properties such as efficient use of contextual coherency within neighboring slices to the input 2D slice and handling inter-subject and intra-subject uptake variability.

3.1. Training Overview

Let $\mathcal{X} = \{X_n \in \mathbb{R}^{h\times w}\}_{n=1}^{N}$ denote a whole-body 3D PET-NC image consisting of N axial slices each of size $h \times w$ degraded by the attenuation function Φ, and $\mathcal{Y} = \{Y_n \in \mathbb{R}^{h\times w}\}_{n=1}^{N}$ be the corresponding reference PET-CT ground truth (section 2). The attenuation degradation function for the n-th slice can be written as:

$$X_n = \Phi(Y_n; \phi); \quad \forall n \in \{1, 2, ..., N\} \tag{1}$$

where ϕ indicates the set of parameters associated with the degradation function. A 2D network for direct attenuation correction aims to determine a predictive function Θ, parameterized by θ, which maps an arbitrary 2D slice in the PET-NC domain, i.e. X_i, to its corresponding slice in PET-AC domain, Y_i. In the conventional 2.5D training scheme, the network takes in a set of M PET-NC slices $\mathcal{X}_m = \{X_i\}_{i-M/2}^{i+M/2}$ centered at i-th slice, known as the target slice, and produces the PET-AC counterpart. Following the formalism of a supervised training scheme and empirical risk minimization, we seek θ^* that minimizes the distance between $\hat{Y}_i$ and Y_i over a training dataset.

3.2. Training Details

We trained our network for 5×10^4 iterations with a batch size of 64. To form the training mini-batches, we firstly selected random subjects and then picked random individual 2D slice from each subject. The selected 2D slices in the mini-batch were subsequently coupled with their immediate previous and next neighbors along the axial dimension. Adam optimizer with default parameters; $\beta_1 = 0.9$ and $\beta_2 = 0.999$ was used for optimization. The learning rate was initially set to 7×10^{-4} and multiplied by 0.1 at iterations 2×10^4 and 3×10^4. The network parameters were initialized with the Kaiming method (He et al., 2015). We set $M = 5$ and optimize the network parameters using mean squared error (MSE).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

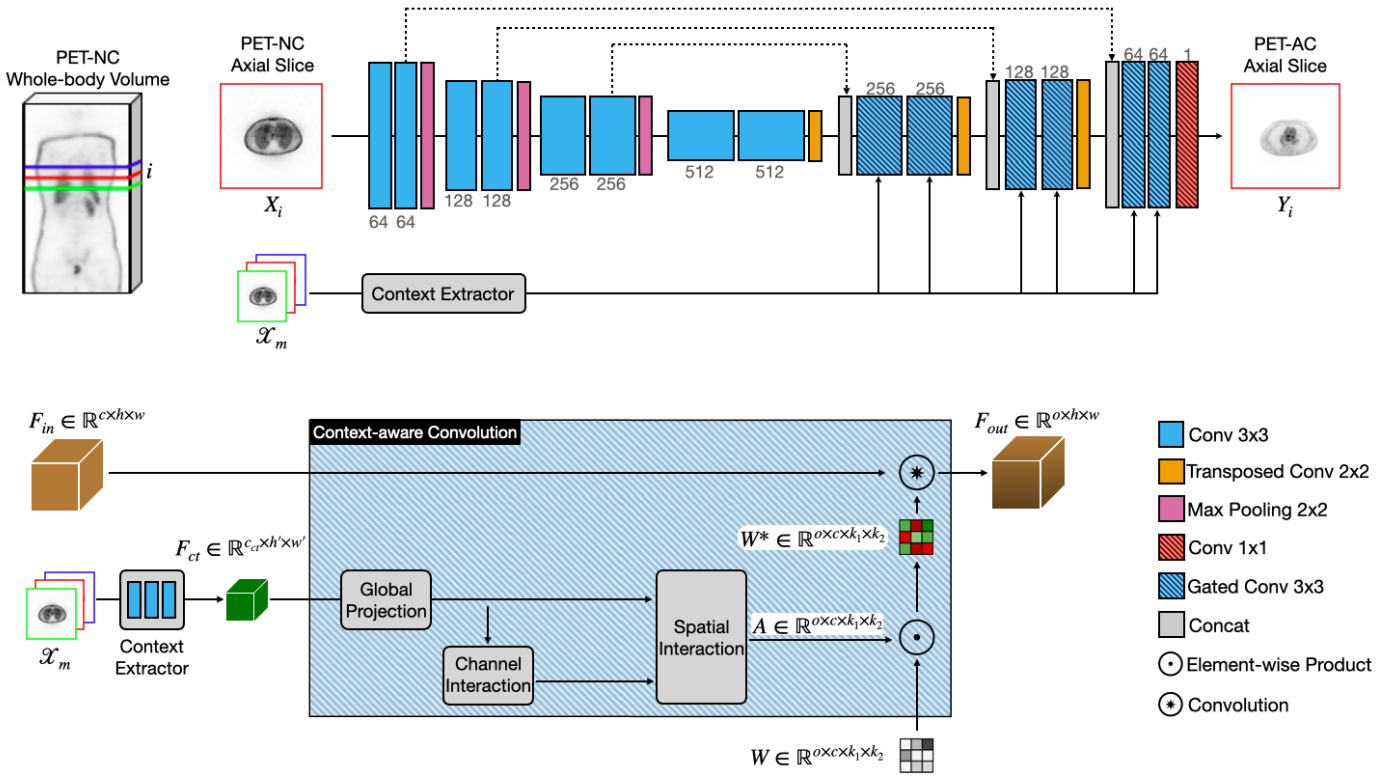


Fig. 1. The overall architecture of our proposed network

3.3. Architecture Overview

Fig. 1 depicts the overview of our proposed CA-DAC network for an efficient attenuation correction. CA-DAC follows the U-Net architecture with $2b$ building blocks. In particular, we employ b blocks in the encoder followed by b symmetric blocks to form the decoder part. Every encoder block consists of two standard convolutional layers with kernel size 3×3 and ReLU activations. Furthermore, 2×2 max-pooling with stride 2 is used after each encoder block to downs-sample the feature maps until a bottleneck layer. For the decoder, transposed convolutional layer is utilized between consecutive blocks to bring the feature maps into the original input resolution. Skip connections link the encoder block to their corresponding decoder block to ensure improved information flow. We replace the standard convolutional layers in the decoder with CAC layers to incorporate the contextual information while preserving the computational efficiency. Each CAC layer takes the 2D input along with its immediate anterior and posterior neighbors, extracts information via a shared sub-network, and modifies the original convolution kernel by attention map computed through a global projection, channel interaction, and spatial interaction. Empirically, we set parameter $b = 5$ and the feature numbers in the decoder to $32, 64, 128, 256$ and 512.

3.4. Context-aware Convolution

In general, the input to a standard convolution is a feature map $F \in \mathbb{R}^{c \times h \times w}$ with c input channels, height h, and width w. Then, a sliding window is swiped over the feature map to extract local patches of size $\mathbb{R}^{c \times k_1 \times k_2}$ with c channels, height h and width w. Next, the patches are convolved with the kernel weights $W \in \mathbb{R}^{o \times c \times k_1 \times k_2}$ to produce the output feature maps with o output channels. As the weights W in standard convolution only depend on the local information and are agnostic to the image context, we propose CAC layers to mitigate the aforementioned issue with standard convolutions. A CAC layer learns a mapping function that outputs a context-aware attention map based on the contextual prior within neighboring input slices. The learned attention map then adaptively influences the convolution kernel in a network designed for direct ASC. Fig 2 provides a detailed illustration of CA layer.

Context Extractor. Here, we focus on the context extraction module. Although several choices can be considered for summarizing the contextual information in neighboring slices, we adopt a shallow network consisting of three 2-strided standard convolutional layers followed by ReLU activation for this purpose (Fig. 1-1). This network takes in the input slices $\mathcal{X}_m$ and extract contextual features $F_{ct} \in \mathbb{R}^{c_{ct} \times h' \times w'}$ where $h' = h/8$ and $w' = w/8$. The extracted feature tensor F_{ct} are then leveraged in different instances of the CAC layers to modify kernel weights based on the contextual information. To improve the efficiency of the network, we keep the depth of the context extractor module to be shallow and share its weights across all instances of the CAC layers.

Global Projection. As the first component in a CAC layer

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

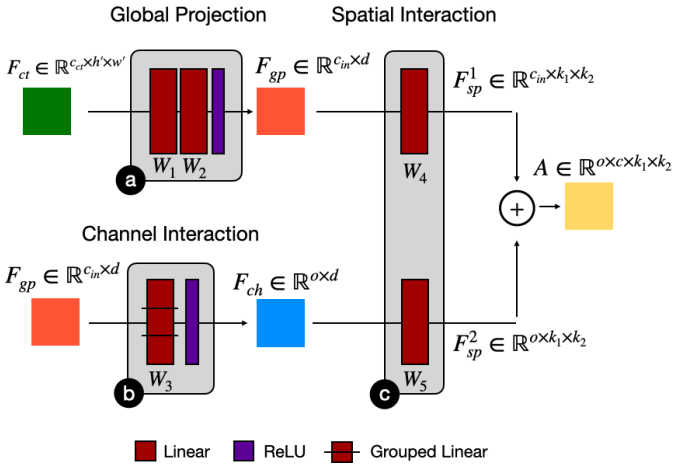


Fig. 2. Detailed illustration of the modules in the context-aware convolutional layer.

(Fig 2-a), we use a linear layer with weights $W_1 \in \mathbb{R}^{c_{ct} \times c_{in}}$ to match the number of channels in F_{ct} to the channels in the incoming features from the previous layer and obtain $F_{ct_{in}} \in \mathbb{R}^{c_{in} \times h' \times w'}$. We then use another linear layer with weights $W_2 \in \mathbb{R}^{h'w' \times d}$ followed by ReLU to project individual flattened channels of size $h'w'$ in $F_{ct_{in}}$ into a global latent representation $F^c_{gp} \in \mathbb{R}^{1 \times d}$. Following the bottleneck design in (Vaswani et al., 2017; Lin et al., 2020; Hu et al., 2018), we set $d = \frac{k_1 \times k_2}{2}$. The weights W_2 are shared across all channels within a CAC layer to reduce the number of learnable parameters. The output for $F_{ct_{in}}$ with c_{in} inputs channels is $F_{gp} \in \mathbb{R}^{c_{in} \times d}$.

Channel Interaction. The second module aims to transform F_{gp} with d channels to the space spanned by the desired number of output channels, i.e. o (Fig 2-b). Following (Lin et al., 2020), we use a grouped linear layer parametrized by $W_3 \in \mathbb{R}^{\frac{d}{g} \times \frac{o}{g}}$ where g is the number of groups. The module ends with a ReLU and gives $F_{ch} \in \mathbb{R}^{o \times d}$ as the output.

Spatial Interaction. The input to the third module is both F_{gp} and F_{ch}, which are decoded to the spatial size of the convolution kernel (Fig 2-c). Similar to (Lin et al., 2020), we use two linear layers with weights $W_4 \in \mathbb{R}^{d \times k_1 \times k_2}$ and $W_5 \in \mathbb{R}^{d \times k_1 \times k_2}$, which are shared across channels in F_{gp} and F_{ch}, respectively. The outputs of the module are $F^1_{sp} \in \mathbb{R}^{o \times k_1 \times k_2}$ and $F^2_{sp} \in \mathbb{R}^{c \times k_1 \times k_2}$. Then, the final attention map is formed by:

$$A_{h,i,j,k} = \sigma\left(F^1_{sp}(i,j,k) + F^2_{sp}(h,j,k)\right) \tag{2}$$

where σ represents the sigmoid function. Given that the original kernel $W \in \mathbb{R}^{o \times c \times k_1 \times k_2}$ and obtained attention mask $A \in \mathbb{R}^{o \times c \times k_1 \times k_2}$ are of the same size, the underlying contextual information within the input slices can be incorporated into the updated kernel W^* through element-wise multiplication, i.e.:

$$W^* = W \odot A. \tag{3}$$

The output feature representations is then obtained by convolving the input feature maps F with the updated kernel W^*.

3.5. Evaluation Strategy

We measured the quantitative error between the reference PET-CT and predicted PET-AC outputs using relative error (RE%) and absolute relative error (ARE%) which are computed as follows:

$$\text{RE\%} = \frac{1}{V}\sum_{i=1}^{V}\frac{\text{PET}_{\text{ASC}}(i) - \text{PET}_{\text{CT}}(i)}{\text{PET}_{\text{CT}}(i)} \times 100\% \tag{4}$$

$$\text{ARE\%} = \frac{1}{V}\sum_{i=1}^{V}\left|\frac{\text{PET}_{\text{ASC}}(i) - \text{PET}_{\text{CT}}(i)}{\text{PET}_{\text{CT}}(i)}\right| \times 100\% \tag{5}$$

where V refers to the total number of voxels, and PET_{ASC} and PET_{CT} indicate the estimated PET-AC images using DL-based approaches and the CT-based ground truth, respectively. In addition to the quantification errors, popular image quality metrics including peak-signal-to-noise-ration (PSNR) and structural similarity (SSIM) were utilized to compare the visual similarity of the estimated PET-AC images against their corresponding PET-CT ground truth images. These metrics are defined as follows:

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}}{\frac{1}{V}\sum_{i=1}^{V}\left(\text{PET}_{\text{ASC}}(i) - \text{PET}_{\text{CT}}(i)\right)^2}\right) \tag{6}$$

$$\text{SSIM} = \frac{(2\mu_{\text{ASC}}\mu_{\text{CT}} + c_1)(2\sigma_{\text{ASC,CT}} + c_2)}{(\mu^2_{\text{ASC}} + \mu^2_{\text{CT}} + c_1)(\sigma^2_{\text{ASC}} + \sigma^2_{\text{CT}} + c_2)}. \tag{7}$$

In Eq. 7, MAX refers to the maximum possible intensity value in the PET-CT image, and μ_{ASC} and μ_{CT} are the mean intensity of the PET-AC obtained by different DL-based ASC approaches and the PET-CT ground truth, respectively. Also, σ_{ASC} and σ_{CT} indicate the standard deviation in the predicted PET and CT-based ground truth images, respectively, and $\sigma_{\text{DL,CT}}$ represents the co-variance between them. Lastly, c_1 and c_2 are constant values to prevent division by very small values (set to default values of $c_1 = 0.01$ and $c_2 = 0.03$ in our experiments).

For statistical analysis, a joint histogram was plotted to show the distribution of the measured uptake correlation between the DL-based predicted PET-AC outputs and the reference PET-CT ground truth averaged over all patients in the test cohort. We carried out all the quantitative assessments as well as the joint histogram within the SUV units (g/mL) in the range of 0.1–20.0. The SUV was calculated as follows: Image-derived uptake [MBq/mL] / injection dose [MBq] $\times$ patient's weight [g].

4. Results

4.1. Quantitative Assessment

The performance evaluation of our proposed context-aware network (CA-DAC) includes validation against the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

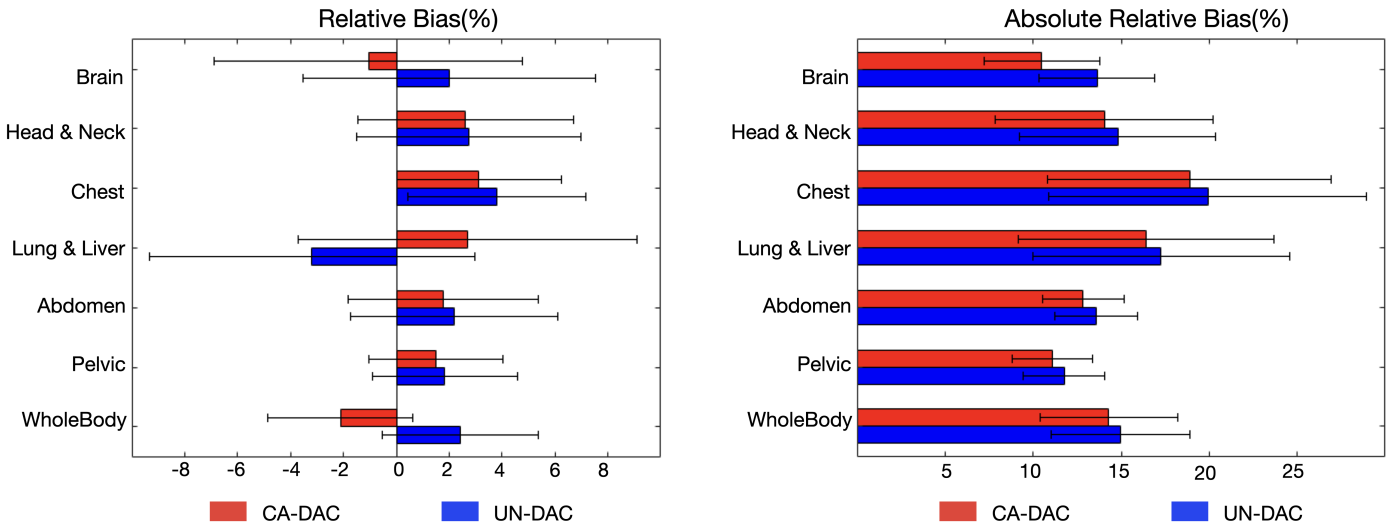


Fig. 3. Quantitative error measured in 6 anatomical regions and whole-body for CA-DAC and UN-DAC approaches over 112 heldout subjects in terms of left) RE(%) and right) ARE(%).

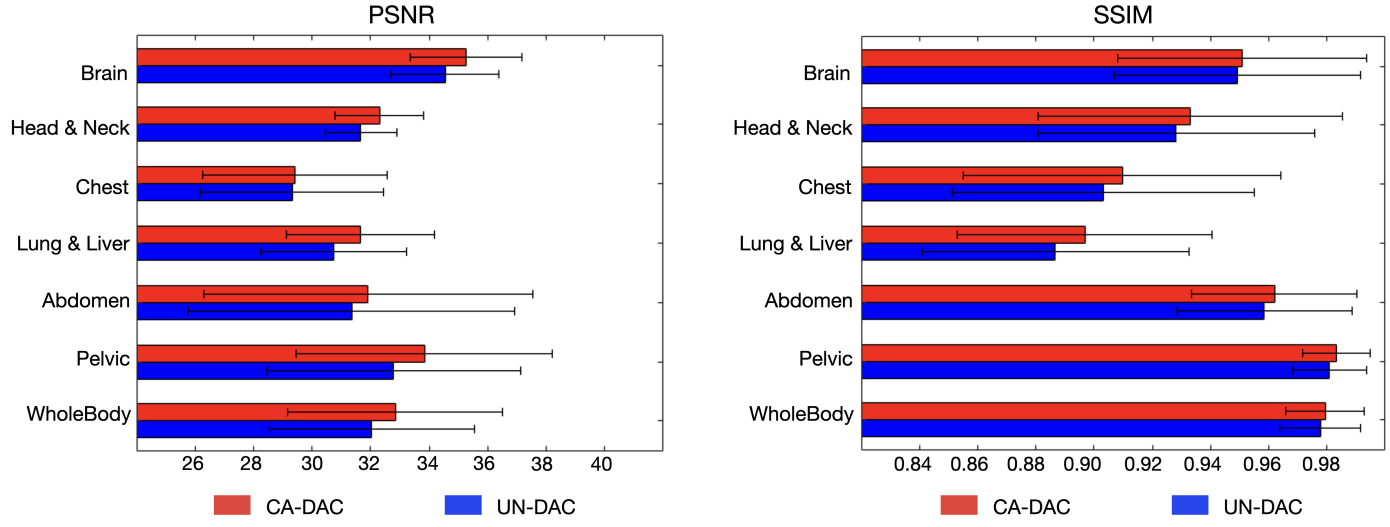


Fig. 4. Quantitative error measured in 6 anatomical regions and whole-body for CA-DAC and UN-DAC approaches over 112 heldout subjects in terms of left) PSNR and right) SSIM.

reference PET-CT ground truth and comparison to a conventional 2.5D UNet (UN-DAC) over a held-out test cohort consisting of 112 subjects across whole-body and 6 anatomical regions. Fig 3(left) depicts the region-wise analysis of PET-AC images in terms of average RE for CA-DAC and UN-DAC. Fig 3(right) further illustrates the region-wise assessment of ARE for CA-DAC and UN-DAC approaches. Our evaluations showed that incorporating the contextual information in the decoder layers results in considerable performance gain across different anatomical regions of the body. In the whole-body, the average RE were 2.43 ± 2.94 and -2.11 ± 2.73 for UN-DAC and CA-DAC, respectively. Also, the average ARE reduced from 14.79 ± 2.37 to 13.96 ± 2.32 for UN-DAC and CA-DAC, respectively. We noted that the highest quantification errors appeared in the chest and lung & liver regions with mean relative error of 3.11 ± 5.11 and 2.68 ± 6.40, respectively. This can be justified by the fact that these regions are often affected by respiratory motion artifacts during the image acquisition and thereby both their appearance and tracer uptake measurements may vary significantly across neighboring slices and ultimately lead to less coherency. The lowest error occurred in the brain region with relative error of -1.05 ± 5.83.

To quantitatively assess the visual quality of the predicted PET-AC images against the reference PET-CT images, Fig. 4 depicts the mean PSNR and mean SSIM in the different regions after attenuation correction by CA-DAC and UN-DAC. From Fig. 3, it is discernible that PET-AC outputs from CA-DAC are visually more similar to the PET-CT ground truth. In particular, we observed that CA-DAC outperformed UN-DAC by 0.71 (dB) averaged over all regions. Likewise, the SSIM scores averaged over all regions were improved from 0.9406 to 0.9465 using CA-DAC

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

over UN-DAC in our evaluation set.

We performed a joint histogram analysis to examine the correlation between the distribution of the corrected SUV values from the predicted PET-AC images and that of the reference PET-CT ground truth. The analysis was carried out within an SUV range of 0.5 to 12 on a log scale. As shown in Fig. 5, CA-DAC yields higher accuracy for lower uptake voxels as the distribution of the correlations exhibits less variation around the identity line. However, the correction accuracy starts to drop for both CA-DAC and UN-DAC in the voxels with higher uptake. The coefficient of determination, denoted by R^2 was further used to quantify the goodness of fit within the joint histograms. In particular, R^2 for CA-DAC and UN-DAC was recorded as 0.982 and 0.963, respectively, indicating a better fit using the former, i.e. proposed method. Moreover, we performed a linear regression analysis by reporting the slope and intercept of the line fitted over the non-zero bins of the joint histograms. The best performance is achieved when the slope and intercept equal to 1.0 and 0.0, respectively. As shown in Fig .5, UN-DAC yielded 0.81 and 1.30 for the slope and intercepts, respectively, while CA-DAC results in 0.88 and 0.57 demonstrating an increase of 0.07 for the slope (closer to 1.0) and a decrease of 0.73 for the intercept (closer to 0.0).

In the next analysis, we studied the impact of context-aware convolution in adapting the behavior of the network to the intra-subject uptake variations. Therefore, we approximately extracted 6 anatomical regions for every subject in the training set using empirical proportional presets over the whole-body image (brain: 10%, head & neck: 15%, chest: 20%, lung & liver: 5%, abdomen: 15%, and pelvic: 35%). We then trained 6 different 2.5D networks for each specific region over the entire training cohort (RGN-DAC) and evaluated them on a region-specific held-out cohort consisting of 50 subjects randomly selected from the primary 112 subjects. These networks provided us with an approximation of the upper-bound performance that could be achieved for each region. We then compared the performance of RGN-DAC against UN-DAC and CA-DAC models trained over whole-body images. As expected, RGN-DAC and UN-DAC yield the lowest and highest absolute relative errors, respectively, as shown in Table 2. However, we observed that context-aware convolutions in CA-DAC successfully enables the the network to outperform UN-DAC by 0.84 in terms of absolute relative error while still lagging behind RGN-DAC by 0.73 averaged over all regions.

Table 3 provides the image quality error computed on whole-body images in three different training schemes, namely UN-DAC (2D) that takes individual slices as input, UN-DAC (3D) which applies 3D convolutions on 3D volumetric patches of size ($64 \times 64 \times 64$), and our proposed CA-DAC, which takes individual 2D slices in the input but adapts the convolution kernels via contextual information. Among the different approaches, UN-DAC (3D) resulted in the largest number of parameters (19M) as it utilized 3D convolutions in the architecture. However, it yielded 0.9792 for SSIM and 32.86 for PSNR, which are not superior than the scores obtained by our proposed CA-DAC, i.e. 0.97936 for SSIM and 32.97 for PSNR. On the other hand, UN-DAC (2D) achieved the lowest scores for SSIM (0.9768), lowest PSNR (32.10), and the smallest number of parameters (7.1M). Compared to UN-DAC, we achieved superior accuracy with negligible increase in model complexity (200K more parameters).

Table 2
PET quantification error measure across 6 regions for UN-DAC, CA-DAC and RGN-DAC

Region	UN-DAC	CA-DAC	RGN-DAC
Brain	18.83 ± 5.92	17.12 ± 4.82	16.23 ± 4.62
Head & Neck	20.73 ± 4.82	20.04 ± 4.22	19.27 ± 5.81
Chest	23.23 ± 6.61	22.92 ± 6.87	22.42 ± 6.34
Lung & Liver	21.92 ± 6.21	21.34 ± 5.69	21.01 ± 5.76
Abdomen	18.48 ± 4.01	17.59 ± 4.81	16.36 ± 4.32
Pelvic	19.66 ± 5.82	18.83 ± 6.11	17.94 ± 5.61

t

Table 3
PET quantification error measured in whole-body for UN-DAC (2D), UN-DAC (3D), and CA-DAC in terms of SSIM and PSNR.

Training Scheme	SSIM	PSNR	# Params
UN-DAC (2D)	0.9768 ± 0.01	32.10 ± 4.52	7.1M
UN-DAC (3D)	0.9791 ± 0.01	32.86 ± 4.01	19M
CA-DAC	0.9793 ± 0.01	32.97 ± 3.79	7.3M

4.2. *Qualitative Evaluation*

For qualitative inspection, we present a comparison of the visual results by UN-DAC and CA-DAC in Fig. 6. It was observed that both of these approaches produced visually indistinguishable results to the naked eyes (but not to SSIM and PSNR) compared to the reference PET-CT. However, as highlighted in section 4, a thorough comparison of their corresponding quantification error maps indicated that UN-DAC over-estimates SUV values in different regions, particularly in the abdomen region. Notably, AC methods usually produce considerable artifacts in the liver dome area due to the local mismatch between the PET and CT images caused by respiratory motion during the scan time. However, the error maps in Fig. 6 showed the our novel CAC layer succeeded to alleviate this problem in this error-prone region compared to UN-DAC. Lastly, Fig. 6 also presents the comparison of the horizontal profile drawn through the abdomen region for PET-CT and the PET-AC images estimated by CA-DAC and UN-DAC. The profile line by CA-DAC accurately matched the one by the reference PET-CT while the one by UN-DAC produced over-estimated SUV measurements for both low and high uptake areas. The qualitative results shown in Fig. 6 were consistent with the quantitative comparisons, which further demonstrates the effectiveness and efficiency of our proposed framework for attenuation correction.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

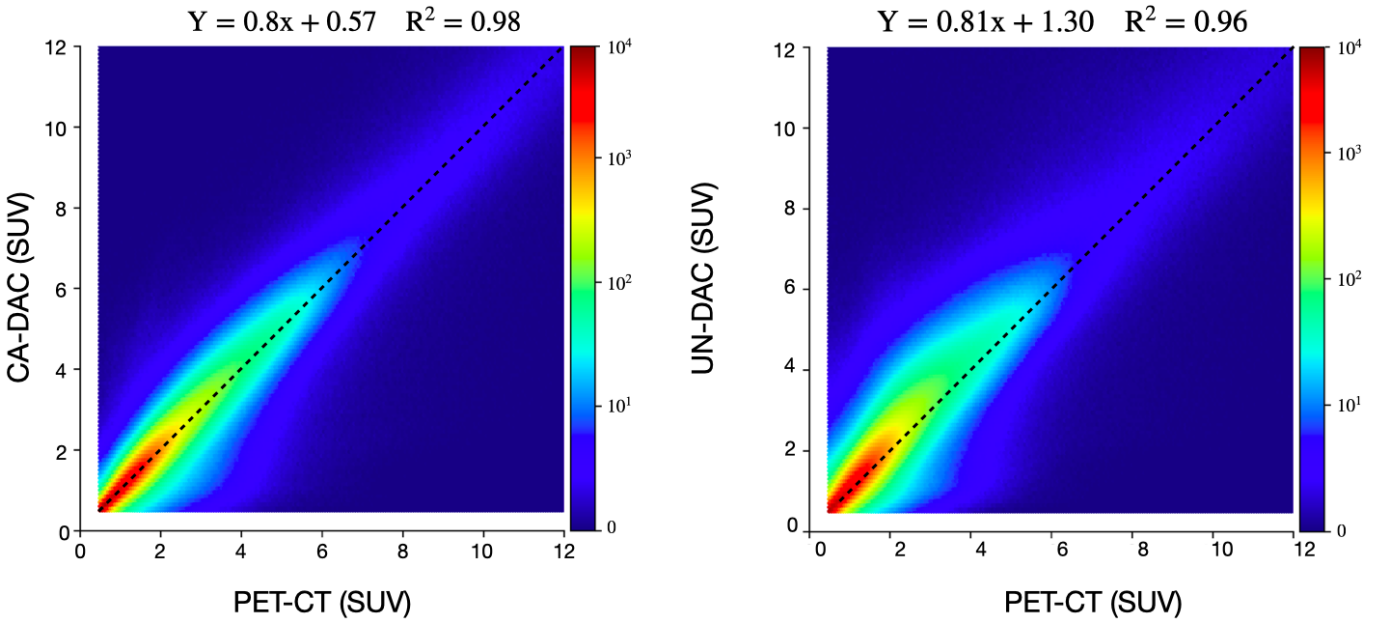


Fig. 5. Joint histogram analysis displaying the correlation between PET-ASC images from left) CA-DAC and right) UN-DAC versus reference PET-CT ground truth. Note that a logarithmic scale was used to display the SUV levels

5. Discussion

In clinical practice, automatic attenuation and scatter correction is an essential yet challenging problem in accurate PET quantification. CT and MRI related artifact can impose significant image quality degradation and/or artifacts on PET images. Artifacts present in anatomical imaging (CT and MRI) may propagate into corresponding PET-AC images. Respiratory mismatches in the thorax between PET and CT/MR images induce banana artifacts in PET images. Metallic and contrast agent material leads to strong streak artefacts and void signal in CT and MRI, respectively which change information of tissue attenuation/scatter factor, which results in inaccurate attenuation/scatter correction of PET images. Truncation artifacts of CT and MRI images appear in patients whose body size exceeds the transaxial field of view of MRI/CT images resulting in erroneous PET quantification. These errors, which propagate to the PET images in the ASC steps of the reconstruction, may not be easily identified due to the absence of a "ground truth". Finally, use of compact PET-only scanners (e.g. for screening purposes) is a possibility in the future if attenuation and scatter correction issues, relying purely on PET images, can be addressed.

Many algorithms, including traditional and deep learning-based methods, have been explored to improve the PET quantification by correcting for attenuation/scatter artifacts. However, these methods failed to reach an efficient trade-off between exploiting the contextual information within neighboring slices and reducing the computational burden. Also, previous methods could not robustly adapt their performance in the presence of intra- and inter-subject uptake variations leading to inferior performance during inference. In this work, we proposed to learn a mapping between individual 2D slices in the PET-NC and PET-AC domains without access to anatomical information provided by either CT or MR. We further suggested augmenting the network architecture with CAC layers which led to a considerable performance gain as evidenced by the quantitative and qualitative results. We further utilized a massive cohort consisting of whole-body images from 910 subjects for training and evaluation.

Some stochastic interaction that causes photons to deviate from a straight line contributes to the attenuation/scatter artifacts. In contrast, the tracer uptakes within adjacent slices are highly correlated and can manifest qualitative and quantitative complementary patterns to each other. Therefore, incorporating the contextual information within neighboring slices into the network can efficiently suppress the impact of random and scatter coincidences recorded during PET image acquisition. In particular, the attenuation and scatter maps inferred in CAC layers can determine how the intermediate representations need to be filtered to accurately restore the PET-AC counterparts. Multiple inputs alone may not suffice to achieve an accurate PET-AC output. Thus, our proposed CA-DAC provides a more flexible ASC algorithm that firstly restores the missing details in every input slice through the inferred dynamic filters and then delivers the final PET-AC estimate for reference slide through eliminating the left over artifact using an averaging operation along slice dimension.

Through extensive quantitative analysis of the obtained PET-AC images (Fig 3 and Fig 4), we showed that our proposed CA-DAC network provided more accurate and robust SUV quantification in different anatomical regions compared to the baseline. The qualitative assessment of the predicted PET-AC images obtained by our proposed CA-DAC (Fig 6) revealed that it could achieve the results with enhanced visual quality, particularly in preserving

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

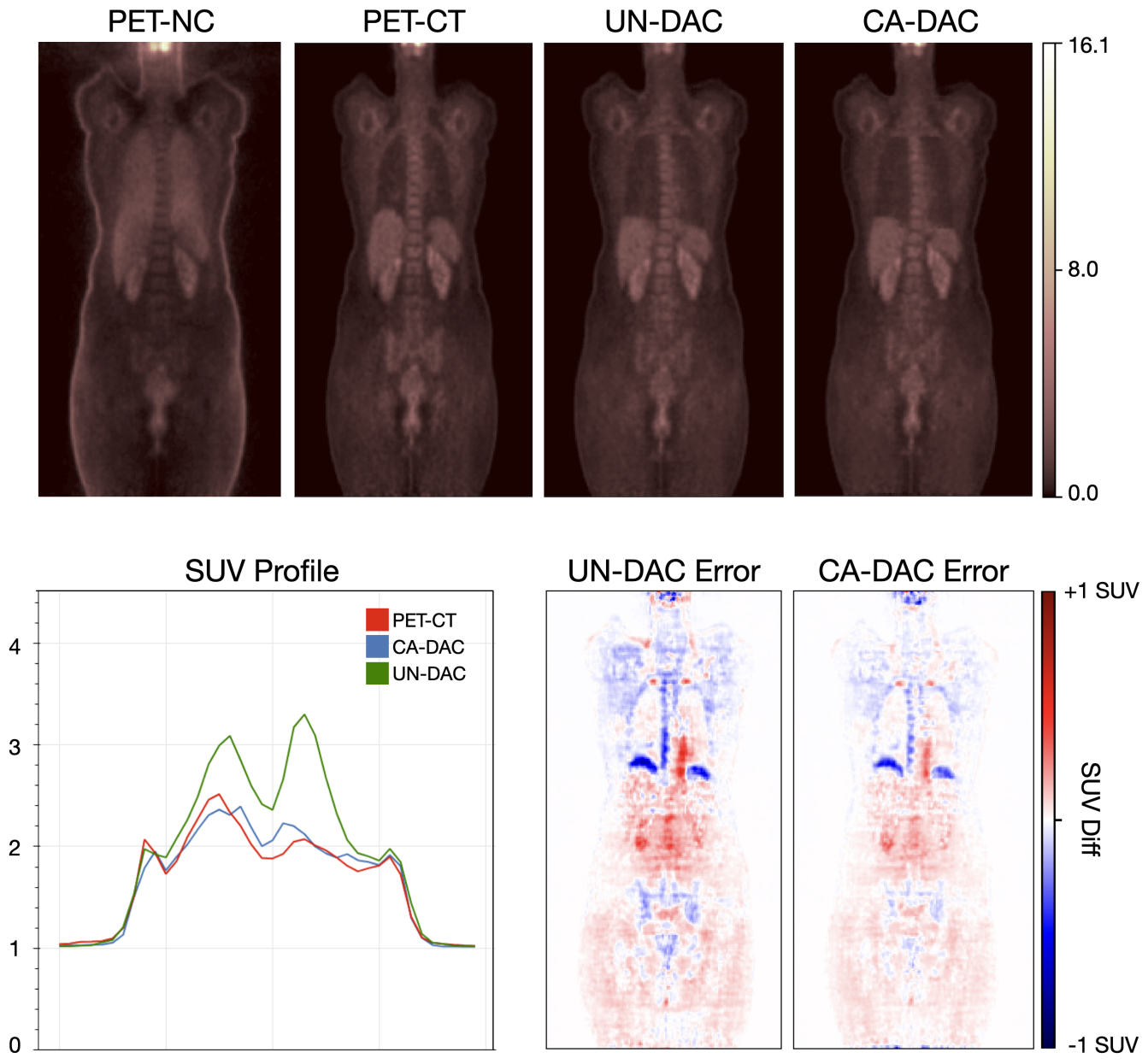


Fig. 6. Qualitative comparison of coronal views of the PET images corrected for attenuation and scatter using UN-DAC and CA-DAC. Top panel provides the PET-NC image, PET-CT ground truth and PET-AC images obtained by UN-DAC and CA-DAC from left to right. Bottom right depicts the error map for UN-DAC and CA-DAC computed against PET-CT ground truths. Bottom left illustrates the tracer uptake profile along a line in abdomen region. A warm color map was chosen to improve the visualization of PET images.

detailed textures and removing sophisticated noise solely from PET-NC inputs. More importantly, visual inspections of the error maps in Fig 6 showed that our proposed CA-DAC could slightly alleviate the quantification error caused by mismatch artifacts in the liver dome region, which is most vulnerable to diaphragm and heart motion. Such improvements are important as it ameliorates the treatment of patients who cannot hold their breath well or in patients who have lesions in the liver dome. In future studies, we will investigate the capability of CA-DAC to address other common artifacts in PET imaging including radio-tracer related (halo artifact), patient related (motion, mismatch,and metal) and instrument related (truncation) artifacts (Shiri et al., 2022). In particular, out attempt will be focused on designing a efficient and fast quality assessment tool with the major objective of detecting and providing quantitative information in regions affected by common artifacts in PET imaging.

6. Conclusion

In this work, we introduced CA-DAC to produce attenuation-corrected PET images without requiring any anatomical information during training and inference. The

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

novelty of the proposed network is to take advantage of context-aware convolutions to modulate the convolution kernels based on the contextual information within neighboring slices along the axial dimension for every 2D input slice. This way, the network can effectively adapt itself to the inter- and intra-subject tracer uptake variations with negligible increase in model complexity. The quantitative and qualitative results indicated considerable performance gain across the whole-body and 6 anatomical regions. This algorithm could be potentially used and evaluated in the context of different PET image artifacts detection and correction including patient and instrument related artifacts.

CRediT authorship contribution statement

Saeed Izadi: Conceptualization, Methodology, Writing - original draft. Isaac Shiri: Data curation, Writing - original draft. Carlos: Data curation. Parham Geramifar: Data curation, Habib Zaidi: Clinical supervision, Writing - original draft, Arman Rahmim: Clinical supervision, Writing - original draft. Ghassan Hamarneh: Technical supervision, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. Acknowledgements

This work was supported by the Swiss National Science Foundation under Grant No. SNSF 320030_176052.

References

Arabi, H., Koutsouvelis, N., Rouzaud, M., Miralbell, R., Zaidi, H., 2016. Atlas-guided generation of pseudo-ct images for mri-only and hybrid pet–mri-guided radiotherapy treatment planning. Physics in Medicine & Biology 61, 6531.

Arabi, H., Rager, O., Alem, A., Varoquaux, A., Becker, M., Zaidi, H., 2015. Clinical assessment of mr-guided 3-class and 4-class attenuation correction in pet/mr. Molecular Imaging and Biology 17, 264–276.

Arabi, H., Zaidi, H., 2020. Deep learning-guided estimation of attenuation correction factors from time-of-flight pet emission data. Medical Image Analysis 64, 101718. URL: http://www.sciencedirect.com/science/article/pii/S1361841520300827, doi:https://doi.org/10.1016/j.media.2020.101718.

Armanious, K., Hepp, T., Küstner, T., Dittmann, H., Nikolaou, K., La Fougère, C., Yang, B., Gatidis, S., 2020a. Independent attenuation correction of whole body [18 f] fdg-pet using a deep learning approach with generative adversarial networks. EJNMMI research 10, 1–9.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020b. Medgan: Medical image translation using gans. Computerized Medical Imaging and Graphics 79, 101684.

Berker, Y., Franke, J., Salomon, A., Palmowski, M., Donker, H.C., Temur, Y., Mottaghy, F.M., Kuhl, C., Izquierdo-Garcia, D., Fayad, Z.A., et al., 2012. Mri-based attenuation correction for hybrid pet/mri systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/dixon mri sequence. Journal of nuclear medicine 53, 796–804.

Berker, Y., Li, Y., 2016. Attenuation correction in emission tomography using the emission data—a review. Medical physics 43, 807–832.

Blanc-Durand, P., Khalife, M., Sgard, B., Kaushik, S., Soret, M., Tiss, A., El Fakhri, G., Habert, M.O., Wiesinger, F., Kas, A., 2019. Attenuation correction using 3d deep convolutional neural network for brain 18f-fdg pet/mr: Comparison with atlas, zte and ct based attenuation correction. PloS one 14.

Boellaard, R., 2009. Standards for pet image acquisition and quantitative data analysis. Journal of nuclear medicine .

Cheuk, D.K., Sabin, N.D., Hossain, M., Wozniak, A., Naik, M., Rodriguez-Galindo, C., Krasin, M.J., Shulkin, B.L., 2012. Pet/ct for staging and follow-up of pediatric nasopharyngeal carcinoma. European journal of nuclear medicine and molecular imaging 39, 1097–1106.

Choi, H., Lee, D.S., 2018. Generation of structural mr images from amyloid pet: application to mr-less quantification. Journal of Nuclear Medicine 59, 1111–1117.

Cohen, J.P., Luck, M., Honari, S., 2018. Distribution matching losses can hallucinate features in medical image translation, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 529–536.

Defrise, M., Rezaei, A., Nuyts, J., 2012. Time-of-flight pet data determine the attenuation sinogram up to a constant. Physics in Medicine & Biology 57, 885.

Dilsizian, V., Bacharach, S.L., Beanlands, R.S., Bergmann, S.R., Delbeke, D., Dorbala, S., Gropler, R.J., Knuuti, J., Schelbert, H.R., Travin, M.I., 2016. Asnc imaging guidelines/snmmi procedure standard for positron emission tomography (pet) nuclear cardiology procedures. Journal of Nuclear Cardiology 23, 1187–1226.

Dong, X., Lei, Y., Wang, T., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2020. Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. Physics in Medicine & Biology 65, 055011.

Dong, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2019. Synthetic ct generation from non-attenuation corrected pet images for whole-body pet imaging. Physics in Medicine & Biology 64, 215016.

Fei, B., Yang, X., Wang, H., 2009. An mri-based attenuation correction method for combined pet/mri applications, in: Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics. p. 726208.

Gong, K., Berg, E., Cherry, S.R., Qi, J., 2019. Machine learning in pet: from photon detection to quantitative image reconstruction. Proceedings of the IEEE 108, 51–68.

Gong, K., Yang, J., Kim, K., El Fakhri, G., Seo, Y., Li, Q., 2018. Attenuation correction for brain pet imaging using deep neural network based on dixon and zte mr images. Physics in Medicine & Biology 63, 125011.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.

Herholz, K., Heiss, W.D., 2004. Positron emission tomography in clinical neurology. Molecular Imaging & Biology 6, 239–269.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

Hwang, D., Kim, K.Y., Kang, S.K., Seo, S., Paeng, J.C., Lee, D.S., Lee, J.S., 2018. Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning. Journal of Nuclear Medicine 59, 1624–1629.

Hwang, D., et al., 2019. 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. Journal of Nuclear Medicine 60, 1183–1189.

Jia, X., De Brabandere, B., Tuytelaars, T., Gool, L.V., 2016. Dynamic filter networks, in: Advances in neural information processing systems, pp. 667–675.

Kinahan, P.E., Townsend, D., Beyer, T., Sashin, D., 1998. Attenuation correction for a combined 3d pet/ct scanner. Medical physics 25, 2046–2053.

Lei, Y., Dong, X., Wang, T., Higgins, K., Liu, T., Curran, W.J., Nye, J.A., Mao, H., Yang, X., 2020a. MRI-aided attenuation correction for PET imaging with deep learning, in: Krol, A., Gimi, B.S. (Eds.), Medical

medRxiv preprint doi: https://doi.org/10.1101/2022.05.26.22275662; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics. SPIE. pp. 521 – 527. URL: https://doi.org/10.1117/12.2549388, doi:10.1117/12.2549388.

Lei, Y., Wang, T., Dong, X., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2020b. Attenuation correction without structural images for pet imaging, in: Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, International Society for Optics and Photonics. p. 113171R.

Leynes, A.P., Yang, J., Wiesinger, F., Kaushik, S.S., Shanbhag, D.D., Seo, Y., Hope, T.A., Larson, P.E., 2018. 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. Journal of Nuclear Medicine 59, 852–858.

Lillington, J., Brusaferri, L., Kläser, K., Shmueli, K., Neji, R., Hutton, B.F., Fraioli, F., Arridge, S., Cardoso, M.J., Ourselin, S., et al., 2020. Pet/mri attenuation estimation in the lung: A review of past, present, and potential techniques. Medical physics 47, 790–811.

Lin, X., Ma, L., Liu, W., Chang, S.F., 2020. Context-gated convolution, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham. pp. 701–718.

Litjens, G., et al., 2017. A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88.

Mehranian, A., Zaidi, H., 2015. Joint estimation of activity and attenuation in whole-body tof pet/mri using constrained gaussian mixture models. IEEE transactions on medical imaging 34, 1808–1821.

Nuyts, J., Dupont, P., Stroobants, S., Benninck, R., Mortelmans, L., Suetens, P., 1999. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. IEEE transactions on medical imaging 18, 393–403.

Rohren, E.M., Turkington, T.G., Coleman, R.E., 2004. Clinical applications of pet in oncology. Radiology 231, 305–332.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Seith, F., Schmidt, H., Gatidis, S., Bezrukov, I., Schraml, C., Pfannenberg, C., la Fougère, C., Nikolaou, K., Schwenzer, N., 2017. Suv-quantification of physiological lung tissue in an integrated pet/mr-system: Impact of lung density and bone tissue. PLOS ONE 12, 1–13. URL: https://doi.org/10.1371/journal.pone.0177856, doi:10.1371/journal.pone.0177856.

Shi, L., Onofrey, J.A., Revilla, E.M., Toyonaga, T., Menard, D., Ankrah, J., Carson, R.E., Liu, C., Lu, Y., 2019. A novel loss function incorporating imaging acquisition physics for pet attenuation map generation using deep learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 723–731.

Shiri, I., Arabi, H., Geramifar, P., Hajianfar, G., Ghafarian, P., Rahmim, A., Ay, M.R., Zaidi, H., 2020. Deep-jasc: joint attenuation and scatter correction in whole-body 18 f-fdg pet using a deep residual network. European Journal of Nuclear Medicine and Molecular Imaging .

Shiri, I., Ghafarian, P., Geramifar, P., Leung, K.H.Y., Ghelichoghli, M., Oveisi, M., Rahmim, A., Ay, M.R., 2019. Direct attenuation correction of brain pet images using only emission data via a deep convolutional encoder-decoder (deep-dac). European radiology 29, 6867–6879.

Shiri, I., Sanaat, A., Salimi, Y., Akhavanallaf, A., Arabi, H., Rahmim, A., Zaidi, H., 2022. Pet-qa-net: Towards routine pet image artifact detection and correction using deep convolutional neural networks, in: IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), IEEE. pp. 1–3.

Su, H., Jampani, V., Sun, D., Gallo, O., Learned-Miller, E., Kautz, J., 2019. Pixel-adaptive convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11166–11175.

Tang, J., Tian, F.P., Feng, W., Li, J., Tan, P., 2020. Learning guided convolutional network for depth completion. IEEE Transactions on Image Processing 30, 1116–1129.

Tao, L., Fisher, J., Anaya, E., Li, X., Levin, C.S., 2020. Pseudo ct image synthesis and bone segmentation from mr images using adversarial networks with residual blocks for mr-based attenuation correction of brain pet data. IEEE Transactions on Radiation and Plasma Medical Sciences .

Vandenberghe, S., Marsden, P.K., 2015. Pet-mri: a review of challenges and solutions in the development of integrated multimodality imaging. Physics in Medicine & Biology 60, R115.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Advances in neural information processing systems, pp. 5998–6008.

Wang, X., Yu, K., Dong, C., Loy, C.C., 2018. Recovering realistic texture in image super-resolution by deep spatial feature transform, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 606–615.

Xu, J., Gong, E., Pauly, J., Zaharchuk, G., 2017. 200x low-dose pet reconstruction using deep learning. arXiv preprint arXiv:1712.04119 .

Yang, X., Lei, Y., Dong, X., Wang, T., Higgins, K., Liu, T., Shim, H., Curran, W., Mao, H., Nye, J., 2019a. Attenuation and scatter correction for whole-body pet using 3d generative adversarial networks. Journal of Nuclear Medicine 60, 174–174.

Yang, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Shim, H., Curran, W.J., Mao, H., Nye, J.A., 2019b. Mri-based attenuation correction for brain pet/mri based on anatomic signature and machine learning. Physics in Medicine & Biology 64, 025001.

Yedder, H.B., Cardoen, B., Hamarneh, G., 2021. Deep learning for biomedical image reconstruction: A survey. Artificial intelligence review 54, 215–251.

Zaidi, H., El Naqa, I., 2021. Quantitative molecular positron emission tomography imaging using advanced deep learning techniques. Annual Review of Biomedical Engineering 23.

Zaidi, H., Koral, K.F., 2004. Scatter modelling and compensation in emission tomography. European journal of nuclear medicine and molecular imaging 31, 761–782.