- ORIGINAL ARTICLE -

Towards a Whole Body [¹⁸F] FDG Positron Emission Tomography Attenuation Correction Map Synthesizing using Deep Neural Networks

Hacia la Sintetización de Mapas de Atenuación de Cuerpo Completo para Tomografía por Emisión de Positrones de [¹⁸F] FDG usando Redes Neuronales Profundas

Ramiro Rodríguez Colmeiro^{1,2,4}, Claudio Verrastro^{2,3}, Daniel Minsky^{3,4}, and Thomas Grosges¹

¹GAMMA3 (UTT-INRIA), Université de Technologie de Troyes, 12 Rue Marie Curie - CS 42060, Troyes cedex, 10004, France ramiro_german.rodriguez_colmeiro@utt.fr

²Universidad Tecnológica Nacional, Sarmiento 440, CABA, C1041AAJ, Argentina ³Comisión Nacional de Energía Atómica, Av. del Libertador 8250, CABA, C1429BNP, Argentina

⁴CONICET, Godoy Cruz 2290, CABA, C1425FQB, Argentina

Abstract

The correction of attenuation effects in Positron Emission Tomography (PET) imaging is fundamental to obtain a correct radiotracer distribution. However direct measurement of this attenuation map is not error-free and normally results in additional ionization radiation dose to the patient. Here, we explore the task of whole body attenuation map generation using 3D deep neural networks. We analyze the advantages that an adversarial training can provide to such models. The networks are trained to learn the mapping from non attenuation corrected $[{}^{18}F]$ -fluorodeoxyglucose PET images to a synthetic Computerized Tomography (sCT) and also to label the input voxel tissue. Then the sCT image is further refined using an adversarial training scheme to recover higher frequency details and lost structures using context information. This work is trained and tested on public available datasets, containing several PET images from different scanners with different radiotracer administration and reconstruction modalities. The network is trained with 108 samples and validated on 10 samples. The sCT generation was tested on 133 samples from 8 distinct datasets. The resulting mean absolute error of the tested networks is 96 ± 20 HU and 103 ± 18 HU with a peak signal to noise ratio of 19.3 ± 1.7 dB and 18.6 ± 1.5 dB, for the base model and adversarial model respectively. The attenuation correction is tested by means of attenuation sinograms, obtaining a line of response attenuation mean error lower than 1% with a standard deviation lower than 8%. The proposed deep learning topologies are capable of generating whole body attenuation maps from uncorrected PET image data. Moreover, the accuracy of both methods holds in the presence of data from multiple sources and modalities and are trained on publicly available datasets. Finally, while the adversarial layer enhances visual appearance of the produced

samples, the 3D U-Net achieves higher metric performance.

Keywords: Attenuation Correction, Deep Learning, Generative Models, Positron Emission Tomography.

Resumen

La corrección de los efectos de la atenuación en las imágenes de Tomografía por Emisión de Positrones (PET) es fundamental para obtener la correcta distribución del radio trazador. Sin embargo la medición directa del mapa de atenuación no está libre de errores y normalmente resulta en la absorción de una dosis superior de radiación ionizante por parte del paciente. Aquí, exploramos la tarea de la generación del mapa de atenuación de cuerpo completo usando resdes neuronales profundas 3D. Se analizan las ventajas que un entrenamiento adversario puede proveer a estos modelos. Las redes son entrenadas para aprender la conversión desde una imagen de $[^{18}F]$ fluorodeoxyglucosa PET sin corrección de atenuación a una imagen sintética de Tomografía Computada (sCT) y además obtener una etiqueta del tipo de tejido en los voxeles de la imagen. Luego la imagen de sCT es refinada usando un entrenamiento de tipo adversario para recobrar detalles de alta frecuencia y estructuras perdidas usando información contextual. Este trabajo es entrenado y probado sobre conjuntos de datos públicos, conteniendo distintas imágenes PET de diferentes tomógrafos, distintos modos de administración de dosis v modos de reconstrucción. La red es entrenada con 108 muestras y validada con 10 muestras. La generación del sCT fue probada con 133 muestras de 8 conjuntos de datos independientes. El error medio absoluto de las redes es de 96 ± 20 HU y 103 ± 18 HU con una relación señal ruido pico de 103 ± 18 HU y 18.6 ± 1.5 dB para el modelo base y el modelo adversario respectivamente. La corrección de atenuación es probada por medio de sinogramas, obteniendo un error medio en la atenuación de las lineas de respuesta menor al 1% con un desvío estándar menor al 8%. Las topologías de aprendizaje profundo propuestas son capaces de generar mapas de atenuación de cuerpo completo a partir de imágenes PET sin corregir. Además, la exactitud de los métodos se sostiene en presencia de datos de múltiples fuentes y modalidades y son entrenadas en conjuntos de datos públicos. Finalmente, mientras se observa que el entrenamiento adversario mejora la apariencia visual de los mapas generados, la topologa 3D U-Net obtiene mejor rendimiento en las métricas.

Palabras claves: Corrección de atenuación, Aprendizaje profundo, Modelos generativos, Tomografía por Emisión de Positrones.

1 Introduction

The correct estimation of attenuation correction maps of positron emission tomography (PET) images is fundamental to their correct reconstruction, but direct measurement of this map means additional ionization radiation dose to the patient. Another approach to obtain this information is to use image analysis methods. These methods create an attenuation structure from other image modality, such as Magnetic Resonance Imaging (MRI) studies or the Non Attenuation Corrected PET (NAC-PET) image. This image translation is specially difficult in whole-body NAC-PET images, since the information it presents is incomplete. In this scenario, where the translation process also needs to fill information blanks, the generative adversarial networks (GANs) are specially powerful.

The application of deep neural models in image to image translation tasks has been successfully exploited in many medical imaging domains, including PET attenuation map synthesizing. However, most methods of attenuation map generation analyze the MRI to CT translation using convolutional networks [1] and GANs with paired [2] and unpaired data [3], requiring a co-registered MRI image wich contains anatomical information that is not present in NAC-PET images. The PET (and NAC-PET) to Computed Tomography (CT) image transalation remains as one of the less explored domains, specially in whole body scans [4]. The studies in this particular domain focus on PET-CT image translation of corrected images on head scans. Liu [5] proposes the use of a 2D U-Net architecture to translate NAC-PET head scans to CT, showing promising results in head region scans. Armanious [2] proposes a general GAN application composed of a cascaded 2D U-Net generator and a discriminator used to evaluate the perceptual loss and style of the generated image. They show the capability of the topology to translate PET scans to CT, using only axial slices and, again, only for head region scans. Both, Liu [5]

and Armanious [2], provide no information on their capability for whole-body image translation which is a harder problem to solve, given that the possible modes in the attenuation structures is larger, including more tissue types (specially soft tissue) and shapes. The whole-body image translation was explored by Dong [6] using a GAN trained with cycle consistency and by Armanious [7] using a 2D GAN based on a cascaded U-Net, however they trained and tested their studies only on state-of-the-art PET scanners with Time of Flight (ToF) cappabilities. The reconstruction of the attenuation map was also studied using maximum-likelihood reconstruction of attenuation and activity (MLAA) methods [8], showing promise in PET scanners with ToF capabilities only when combined with neural networks [9]. However without using a post-processing step MLAA outputs noisy attenuation maps, with or without ToF information [10]. Solving the problem of generating an attenuation map directly from image data can enable non-ToF scanners to use synthetic attenuation maps.

Here we analyze the aptitude of two methods to generate high quality whole body attenuation sinograms by means of artificial CT images from NAC-PET images, a 3D U-Net and a fully 3D GAN topology with mixed loss. Given that, the dimensionality of the 3D volumes is comparable to the generation of high resolution 2D images. Therefore we perform a two step training, starting with supervised training of labels and then we add an adversarial loss block to enhance the image resolution. Our models are trained on a public available data-set from the Cancer Image Archive [11], the data-set contains series of co-registered (adquired with PET-CT scans) CT, PET and NAC-PET whole body scans of Head and Neck Scamorous Cell Carcinoma (HNSCC) [12], acquired with Discovery ST/STE/RX General Electric scanners. We use 8 different datasets for the testing process, containing 5 different types of carcinomas and multiple scanners models: Discovery ST/STE/RX/LS/IQ/610/690, from General Electric and Biograph from Siemens. Only the Discovery 690 possess ToF capabilities.

2 Material, methods and theory

2.1 Topology Description

Two different architectures are tested in this work. First a baseline 3D U-Net topology, trained in a fully supervised manner. Second, a GAN whose generator is composed of the baseline 3D U-Net with additional layers trained in adversarial manner against a convolutional critic (or discriminator). In order to reduce the adversarial training instability, the adversarial gradient flow in the generator does not flow into the baseline 3D U-Net.

2.1.1 Generator:

The model representation can be seen in Fig. 1. The initial section of the generator is a 3D U-Net topology [13] (the 3D U-Net block). The U-Net network posses 5 resolution levels, each of them composed of two convolutional layers with filter shape of $3 \times 3 \times 3$ and Rectified Linear Unit (ReLU) activation. Each resolution level posses a skip connection between the down-sample and up-sample path. Instead of convolutional resampling the resolution changes are performed using trilinear up- or down-sampling. After each convolutional layer we apply a voxel normalization along feature maps, dividing each voxel value by $\sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(vx_j^i)^2 + e}$, where *N* is the number of channels in the feature map, vx_j^i is the voxel value of the i^{th} channel of the j^{th} voxel and $e = 1.0 \times 10^{-8}$. We also apply at each convolutional layer, a scaling factor to the filter kernel based on He's [14] scaled initialization of weights. After the 3D U-Net block the network forks into two branches, inside the Auxiliary Task block. The first branch is used for segmentation, it is composed of three convolutional layers and ends in a softmax layer. The second branch is responsible for the synthetic CT (sCT) generation, it is composed of a convolutional layer with an hyperbolic tangent activation. Finally, the outputs of the 3D U-Net are merged and processed by the GAN layers in the GAN Refinement block. This last block, a collection of 5 convolutional layers with 8 filters each, is used during the adversarial training. All convolutional operations use a filter of size $3 \times 3 \times 3$ except the output layer which use a $1 \times 1 \times 1$ filter.

2.1.2 Critic:

The critic or discriminator network is a convolutional network with ReLU activation in all layers, only the last layer has no activation. The input of this network is a two channel volume composed of the NAC-PET volume and the real or sCT image. The output of the network is a value proportional to quality value of the generated image. The network is conformed by 4 resolution levels with two convolutional layers per level. Each convolution has a filter size of $3 \times 3 \times 3$ and ReLU activation. No batch or pixel normalization is applied. The last two layers of the critic are a flatten operation followed by a single dense layer with linear output.

2.2 Training Scheme

The training of the network is divided into two stages: first the generator network is trained in a supervised manner using a composed loss. The segmentation branch of the network applies a 3D-DICE as shown in Eq. 1,

$$L_D = \frac{1}{N_c} \sum_{c=0}^{N_c} \frac{2\sum_i^{N_v} p_{i,c} g_{i,c}}{\sum_i^{N_v} p_{i,c}^2 + \sum_i^{N_v} g_{i,c}^2},$$
 (1)

where N_c is the number of objective classes, N_v the number of voxels in the volume, $g_{i,c}$ are the voxels of the ground truth and $p_{i,c}$ the values of the softmaxed output of the network. The dice loss ranges from 0 to 1. It produces its maximum value when all voxels of the ground truth $(g_{i,c})$ have the same value as the softmaxed output voxels $(p_{i,c})$. Since the output is softmaxed the denominator of Eq. 1 is always larger than its numerator except when $g_{i,c}$ and $p_{i,c}$ are identical. In the case of a multi-class problem $(N_c > 1)$ the final value is divided by the number of classes.

The CT synthetization branch, up to the GAN layers, is trained using as loss function the euclidean distance between the sCT and the objective CT image, $L_2 = ||s,r||^2$, where s is the sCT and r is the real CT volume. We chose the L_2 over other metrics such as the L_1 metric (which is known to provide a sharper output images) since the L_2 severely penalizes the data outliers. In the case of CT images these outliers correspond to high attenuation structures such as specific parts of the bone tissue. Since for the same percentage variation the L_2 produces larger values when compared to the L_1 , the L_2 metric is assigning more importance to higher attenuation zones which is desirable in this specific problem.

The loss for the supervised training stage is shown in Eq. 2,

$$L_{gen}^{sup.} = (1 - L_D) + K_e L_2,$$
(2)

where K_e is a coupling constant. After the initial training, the adversarial training starts. The adversarial training uses the Wasserstein-GAN (W-GAN) strategy [15], resulting in a generator loss as shown in Eq. 3,

$$L_W = -f_c(f_g(x)), \tag{3}$$

where f_c is the critic network function, f_g is the generator network function and x is the input NAC-PET image. During the adversarial training the GAN layers become active and are trained using the W-GAN loss. The gradient of the GAN does not flow into the 3D U-Net layers. The critic is trained using a coupled pairs of NAC-PET and CT images, real or fake. It is trained using the wasserstein loss shown in Eq. 4,

$$L_{crit.}^{GAN} = f_c(r) - f_c(f_g(x)) + \lambda G_p(s, r), \qquad (4)$$

where G_p is the gradient penalty [16] and $\lambda = 10.0$. The critic is trained for 5 steps for each generator step. At the initial step of the GAN training stage the critic is optimally trained before starting the GAN training loop. The generators are trained using an Adaptive Moment Estimation (ADAM) optimizer with parameters $\beta_1 = 0.0$, $\beta_2 = 0.99$ and $\varepsilon = 1.0 \times 10^8$ and learning rate lr = 0.0001. The discriminator uses a RMSprop optimizer with learning rate lr = 0.0005.



Figure 1: Generator network topology based on the 3D U-Net. Top branch in the *auxiliary task* block is used for label segmentation and bottom branch for artificial CT generation. The output of the *GAN refinement* block is the synthetic CT refined by the GAN layers and built from the *auxiliary task* block outputs.

2.3 Train Dataset Description

The HNSCC dataset is composed of a series of coregistered CT and NAC-PET scans of head and neck squamous cell carcinoma, acquired with PET-CT scanners. However some samples are incomplete or contain only fragments of the body. The selected samples of the dataset consist of whole-body studies. The dataset is first stripped of all samples that do not contain NAC-PET and CT matched samples. Then the matched samples are tested for FoV overlapping and cropped to contain only axial slices with information from all the image types. After the dataset cleaning process is finished, the image size is normalized to a $128 \times 128 \times 256$ voxels FoV, with a voxel size of $5.46 \times 5.46 \times 5.08$ mm³. The final dataset contains 118 images from 71 different patients, from which 7 patients (10 images) were separated as validation dataset. Before feeding each sample to the network, the volume is randomly sliced into a $128 \times 128 \times 32$ volume and all the input NAC-PET voxels values are added the same random value of $\pm 10\%$ of full scale and re-normalized.

2.3.1 Objective CT normalization:

The objective CT must be free of the couch structures in order to train a NN that is independent of the scanner model. A specific couch can be added later to the generated CT when it is used for correction in a given scanner model. The removal of the couch is performed using a method based on the voxel variance along the axial axis [17]. Then the image dynamic range is clipped between -125 and 1300 Housfield Units (HU) and normalized between 0 and 1, to maximize the distance between soft and bone tissue.

2.3.2 Label Generation:

Four labels classes are extracted from the nonnormalized couch stripped CT image using a voxel value threshold. The *Air-Lung* mask values ranges from -1000 HU to -125 HU, the *Fluids-Fat* mask ranges from -125 HU to 10 HU, the *soft-tissue* mask ranges from 10 HU to 90 HU and the *Bone* mask ranges from 90 HU to 1300 HU.

2.4 Test Dataset Description

The test datasets are series of public dataset, also from TCIA, including different types of lesions, patients and scanner technologies. The dataset is composed of 133 test samples: 73 from the Non-Small Cell Lung Cancer (NSCLC) [18] dataset, 25 from The Cancer Genome Atlas - Head-Neck Squamous Cell Carcinoma (TCGA-HNSC) [19], 20 from The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) [20], 1 from The Cancer Genome Atlas - Thyroid Cancer (TCGA-THCA) [21], 4 from Clinical Proteomic Tumor Analysis Consortium - Lung Adenocarcinoma (CPTAC-LUAD) [22], 1 from Clinical Proteomic Tumor Analysis Consortium - Pancreatic Ductal Adenocarcinoma (CPTAC-PDA) [23], 3 from Clinical Proteomic Tumor Analysis Consortium - Uterine Corpus Endometrial Carcinoma (CPTAC-UCEC) [24] and 6 from Clinical Proteomic Tumor Analysis Consortium - Lung Squamous Cell Carcinoma (CPTAC-LSCC)[25]. These datasets were cleaned from non-matching samples and

re-normalized to fit voxel size expected by the network. During the testing procedure the NAC-PET is fed to the model in consecutive slices of $128 \times 128 \times 32$ voxels. The intensity is re-scaled to [0,1] for each slice (no other intensity normalization is applied). Since the output of the network is not a full FoV, the consecutive outputs of the network are composed into a single volume. This composition is done by multiplying the current slice with a Gaussian bell along the z-axis. The Gaussian bell is centered in the middle of the current slice and has a $\sigma = 4$ voxels. After, the weighted slice is added to the full FoV at its corresponding location. Finally the FoV is divided voxelwise by the sum of the weights of slices that provided information to each voxel, in order to obtain the weighted mean voxel value.

2.5 Attenuation Correction Metrics

Since the sinogram data of the used datasets is not available, direct reconstruction of the corrected PET activity using the proposed model is not possible. As a quantitative assessment of the attenuation correction performance we compared the Line of Response (LoR) attenuation using the provided CT and the generated sCT. The LoR attenuation was constructed using attenuation sinograms with size (180×180) and an axial step of 5 mm. Each CT and sCT was first converted from Hounsfield Unit scale to linear attenuation $[1/cm^2]$, at 120 keV. Then the linear attenuation was scaled to the PET energy, 511 keV. Using the real CT and the sCT, a pair of attenuation sinograms was created. For each of the LoRs in the sinograms, the attenuation coefficient was calculated as $At_i = e^{-\int \Delta M(s_i) ds_i}$, where s_i represents the i^{th} LoR of the sinogram. Using the provided CT as ground truth, the differences in the attenuation sinograms was measured using the mean difference and standard deviation of the sinogram values. This was applied in four samples, the HNSCC-01-0148 sample from the validation dataset, sample AMC-009 from NSCLC Radiogenomics, C3N-00957 from CPTAC-PDA and TCGA-BB-7863 from TCGA-HNSC.

2.6 Image Quality Metrics

The generated sCT image quality was tested against the ground truth CT using three different metrics, Peak Signal to Noise Ratio (PSNR), Mean Absolute Error (MAE) and Normalized Cross Correlation (NCC).

2.7 Ablation Tests

In order to assess the importance of the components in the presented methods, an ablation test is performed. Specifically, the presence of the segmentation branch and the restriction of the GAN gradient are analyzed. The test is applied on the image quality metrics for: a 3D U-Net GAN trained using only the traditional loss, described in Eq. 3 and without restricting the gradient flow, named *Base GAN*; a 3D U-Net with no segmentation operation, named *No Seg. U-Net*; a 3D U-Net GAN with no segmentation information, named *No Seg. GAN*. The ablation tests results are shown along the metrics of the *U-Net* and *GAN* networks. All networks were trained for the same number of steps.

2.8 Reconstruction Tests

A reconstruction quantitative test of the correction capabilities of the sCT was constructed by projecting the corrected activity image provided in the dataset. This process is done by taking an attenuation corrected sample and its corresponding attenuation map and performing a large number of randomly sampled 3D projection operations (more than 1×10^{10} LoRs). Then the projections are reconstructed using a Back-Projection Filtered (BPF) algorithm. The algorithm was chosen to avoid storing a large number of projections. It is worth noting that this reconstruction process produces lower quality samples when compared to fully iterative processes such as Maximum Likelihood Expectation Maximization (MLEM). This test was applied to two samples: the AMC-009 sample from the NSCLC Radiogenomics test dataset and the C3N-00957 from the CPTAC-PDA test dataset.

3 Results

3.1 Attenuation Correction

The resulting metrics of the attenuation sinograms are summarized in table 1. A central ring attenuation sinogram of the reference CT and the sCT are displayed in Fig. 2 for the validation sample HNSCC-01-0148 and in Fig. 3 for the AMC-009 sample from NSCLC Radiogenomics dataset. The sinograms are displayed along the histogram of the difference values for all the sinograms in the FoV, showing the shape of the error distribution presented in table 1.

3.2 Image Quality

The PSNR, MAE and NCC metrics of the generated sCT to the ground truth CT are presented in the Fig. 4, each of the box plots corresponds to the metrics of the 3D U-Net and GAN on the different test sets. The datasets are presented by source since the number of samples of the individual datasets is too small in some cases. The scores for the whole validation and test dataset are summarized in table 2.

Three samples from the test datasets can be seen in Fig. 5, Fig. 6, and Fig. 7, two with the patients arms elevated over the head (arms up) and other with the patients arms positioned at the side (arms down). These images correspond to the NSCLC Radiogenomics, CPTAC-PDA and TCGA-HNSC datasets respectively. In these figures the sCT images, generated using only

Table 1: Mean and standard deviation of the difference between reference CT attenuation and sCT attenuation, in percentage of the reference value, for a validation sample and three test samples. A good metric will be close to zero and have a low standard deviation.

Topology	HNSCC-01-0148	AMC-009	C3N-00957	TCGA-BB-7863
U-Net	$-0.43 \pm 4.0\%$	$-0.53 \pm 5.5\%$	$0.34 \pm 6.3\%$	$-0.49 \pm 3.6\%$
GAN	$-0.71 \pm 4.2\%$	$-0.65 \pm 5.8\%$	$0.41 \pm 7.7\%$	$-0.66 \pm 3.9\%$



Figure 2: Central ring from the attenuation sinograms from the 3D-Unet sCT (a), reference CT (b) and the GAN sCT (c) of the HNSCC-01-0148 sample from the validation dataset. Below, the histograms (d,e) of the LoR attenuation value differences, of the 3D-Unet sCT and the 3D GAN sCT respectively.



Figure 3: Central ring from the attenuation sinograms from the 3D-Unet sCT (a), reference CT (b) and the GAN sCT (c) of the AMC-009 sample from NSCLC Radiogenomics dataset. Below, the histograms (d,e) of the LoR attenuation value differences, of the 3D-Unet sCT and the 3D GAN sCT respectively.

Table 2: Validation (10 samples) and testing (133 samples) scores for the proposed 3D U-Net and GAN topology (bold font). The ablation test results are also included, the *No Seg.* prefix means that the segmentation operation was not implemented.

Data			r 7
Data	PSNR [dB]	MAE [HU]	NCC $[-]$
Val	17.9 ± 1.8	120 ± 25	0.743 ± 0.060
Test	17.7 ± 1.6	124 ± 23	0.610 ± 0.095
Val.	21.0 ± 1.4	80 ± 13	0.802 ± 0.050
Test	19.3 ± 1.7	96 ± 20	0.760 ± 0.060
Val	18.3 ± 1.3	121 ± 20	0.640 ± 0.080
Test	17.3 ± 1.5	133 ± 22	0.600 ± 0.070
Val	18.6 ± 1.6	118 ± 25	0.63 ± 0.11
Test	17.5 ± 1.4	128 ± 20	0.585 ± 0.085
Val.	19.9 ± 1.3	89 ± 10	0.760 ± 0.050
Test	18.6 ± 1.5	103 ± 18	0.720 ± 0.060
	Val Test Val. Test Val Test Val Test Val. Test	Val 17.9 ± 1.8 Test 17.7 ± 1.6 Val. 21.0 ± 1.4 Test 19.3 ± 1.7 Val 18.3 ± 1.3 Test 17.3 ± 1.5 Val 18.6 ± 1.6 Test 17.5 ± 1.4 Val. 19.9 ± 1.3 Test 18.6 ± 1.5	Val 17.9 ± 1.8 120 ± 25 Test 17.9 ± 1.8 120 ± 25 Test 17.7 ± 1.6 124 ± 23 Val. 21.0 ± 1.4 80 ± 13 Test 19.3 ± 1.7 96 ± 20 Val 18.3 ± 1.3 121 ± 20 Test 17.3 ± 1.5 133 ± 22 Val 18.6 ± 1.6 118 ± 25 Test 17.5 ± 1.4 128 ± 20 Val. 19.9 ± 1.3 89 ± 10 Test 18.6 ± 1.5 103 ± 18

the supervised loss shown in Eq. 2, and the sCT images, generated using the adversarial loss shown in Eq. 3, are compared against the reference CT. Also these figures present the 3D views of the skeletal tissue generated by each network topology and the reference structure.

3.3 Reconstruction Tests

The resulting tracer distribution obtained using the real CT and each sCT are shown in Fig. 8 for the test sample AMC-009 from the NSCLC Radiogenomics dataset and in Fig. 10 for the test sample C3N-00957 from CPTAC-PDA dataset. In Figs. 9 and 11 the activity profiles along three different zones (marked by a dashed line in Figs. 8 and 10) are shown for the AMC-009 and C3N-00957 sample respectively.

4 Discussion

The attenuation correction capability of the networks were tested using a series of attenuation sinograms and comparing the sCT sinograms to the real CT. This metric achieved a mean error lower than 1% and a standard deviation below 8% in the worst case, as seen in Table 1, which can be improved using a larger and more uniform datasets (in terms of anatomical coverage). While the addition of GAN layers improves the visual appearance of the sCT, an improvement in the



Figure 4: Boxplot of PSNR (a,d), MAE (b,e) and NCC (c,f) for each test data source and validation dataset. The figures (a,b,c) and (d,e,f) are corresponding to the 3D U-Net and to the network with adversarial refinement, respectively.



Figure 5: Input NAC-PET (a,g), 3D-Unet sCT (b,e,h), 3D GAN sCT (c,f,i) and reference CT (d,j,k) coronal (a,b,c,d), sagittal (g,h,i,j) and 3D renders of bone tissue (e,f,k) of AMC-009 from NSCLC Radiogenomics. The improvement of the GAN can be seen in the enhancement of higher frequency details between images (b, c) and (h, i). In the 3D render, the GAN network (f) is able to generate further details than the 3D U-Net (e), when compared to the ground truth (k).



Figure 6: Input NAC-PET (a,g), 3D-Unet sCT (b,e,h), 3D GAN sCT (c,f,i) and reference CT (d,j,k) coronal (a,b,c,d), sagittal (g,h,i,j) and 3D renders of bone tissue (e,f,k) of C3N-00957 from CPTAC-PDA dataset.



Figure 7: Input NAC-PET (a,g), 3D-Unet sCT (b,e,h), 3D GAN sCT (c,f,i) and reference CT (d,j,k) coronal (a,b,c,d), sagittal (g,h,i,j) and 3D renders of bone tissue (e,f,k) of TCGA-BB-7863 from TCGA-HNSC dataset.



Figure 8: Reconstructed NAC-PET (a,g), PET image corrected using the reference CT (b,h), PET image corrected using 3D-Unet sCT (c,i), difference map of PET corrected with the reference CT and the corrected with the 3D-Unet sCT (d, j), PET image corrected using 3D GAN sCT (e,k) difference map of PET corrected with the reference CT and the corrected with the GAN sCT (f, i), coronal cuts (a, b, c, d, e, f) and axial cuts (g, h, i, j, k, l) of sample AMC-009 from NSCLC Radiogenomics dataset. The red dashed lines mark the locations of the profiles shown in Fig. 9. The intensity in images (a) and (g) is out of scale in order to illustrate the difference to the corresponding attenuation corrected images. The attenuation correction is noted in images (b, c, e) and (h, i, k) as a reduction in the relative border to center intensity presented in images (a) and (g), respectively.



Figure 9: Activity profiles of the PET reconstruction profile using the reference CT (black dashed line), using the GAN sCT (solid red line), using 3D U-Net sCT (solid green line) and without correction (dashed cyan line) along the red dashed lines presented in Fig. 8 of sample AMC-009 from NSCLC Radiogenomics dataset. The NAC PET profile is out of scale in order to illustrate the shape difference to the corresponding attenuation corrected profiles. The differences observed in the left side of (a) are corresponding to respiration effects on the base of the lung. The differences observed near the center of (b,c) are corresponding to a hip prosthesis (also observable in Fig. 5(d)).



Figure 10: Reconstructed NAC-PET (a,g), PET image corrected using the reference CT (b,h), PET image corrected using 3D-Unet sCT (c,i), difference map of PET corrected with the reference CT and the corrected with the 3D-Unet sCT (d, j), PET image corrected using 3D GAN sCT (e,k) difference map of PET corrected with the reference CT and the corrected with the GAN sCT (f, i), coronal cuts (a, b, c, d, e, f) and axial cuts (g, h, i, j, k, l) of sample C3N-00957 from the CPTAC-PDA dataset. The red dashed lines mark the locations of the profiles shown in Fig. 11. The intensity in images (a) and (g) is out of scale in order to illustrate the difference to the corresponding attenuation corrected images. The attenuation correction is noted in images (b, c, e) and (h, i, k) as a reduction in the relative border to center intensity presented in images (a) and (g), respectively.



Figure 11: Activity profiles of the PET reconstruction profile using the reference CT (black dashed line), using the GAN sCT (solid red line), using 3D U-Net sCT (solid green line) and without correction (dashed cyan line) along the red dashed lines presented in Fig. 10 of sample C3N-00957 from the CPTAC-PDA dataset. The NAC PET profile is out of scale in order to illustrate the shape difference to the corresponding attenuation corrected profiles. The differences observed in the center of (a) are corresponding to respiration effects on the base of the lung.

attenuation correction is not observed. Contrary to the expected result, the 3D U-Net model quality seems to be more adequate. A possible explanation of this could be found in the fact that the GAN layers are trained using the gradient of the critic network. When compared to a specific metric, such as the L_2 , the critic responds differently. While the L_2 provides a value directly associated to the difference from a ground truth at a voxel scale, the critic provides a value that roots in an internal representation how the real sample should be. The critic is trained to learn an internal representation of the set of real CT images and it is conditioned to a given input, a NAC PET in this case. The specificity of this internal representation depends largely on the critic network capacity, how representative is the dataset and the training state of the critic (how good is it at detecting an sCT). Given that the capacity of the critic (and also the generator) was constrained by the available computing power and that the dataset cannot be considered infinite for the size of this problem, we believe that the representation power of the critic network was sufficient to produce better looking images but lacks specificity to internalize other aspects that are important for the PET image correction task. It should be possible to improve the results by increasing the size of the critic (and generator networks) and/or by increasing the amount of training samples.

The tested models show to be resistant to multiple reconstruction techniques and scanners technologies when operating in the selected image resolution, as shown by the test metrics in Fig. 4. The basic 3D U-Net topology generates synthetic attenuation correction images with a PSNR of 19.3 ± 1.7 dB, a MAE of 97 ± 20 HU and NCC of 0.760 ± 0.064 . The addition of the GAN layers achieve a PSNR of 18.6 ± 1.4 dB. a MAE of 103 ± 18 HU and NCC of 0.720 ± 0.059 . These scores are obtained on test samples from different scanners, patients and lesions, showing that these techniques can be used on multiple sources. It can be seen that the 3D U-Net network generates images dominated by lower frequency components whereas the GAN trained network shows higher frequency details and borders, as shown in Fig. 5. Nevertheless these improvements are not reflected in the performance of the metrics. The difference between both networks was assessed with a t-test using a significance level of $\alpha = 0.05$. The obtained *p*-values are: $p_{PSNR} = 0.00015, p_{MAE} = 0.0036, p_{NCC} = 1.2 \times 10^{-7}$ all below the significance level, resulting in statistically meaningful differences.

Both, the 3D U-Net and the adversarial networks fail to generalize the upper section of the body where less training data were available. This is also reflected in the variability in the arm's postures in the dataset. While this can be solved with additional training data, it can also be mitigated by anatomically matching the training data and train region specific networks. Further improvement in this direction will be to train the network using a full-size intermediate space to map each anatomical section, such as the intermediate representations presented in [26] and [27]. The current sCT generation can also be used as prior on attenuation reconstruction techniques such as the MLAA and single scatter modeling [28], reducing their high computational cost. These techniques can potentially eliminate artifacts from the generated attenuation maps, such as the CT contrast in the stomach observed in Fig. 7(d) that is not present in Fig 7(b) nor in Fig 7(c), however it can be noted in the NAC PET image in Fig 7(a), as a darkened region.

The importance of two of the main features of the tested topologies was assessed using an ablation test. It can be seen in table 2 that the inclusion of the segmentation branch as an auxiliary task improves significantly the metric scores when trained for the same number of steps ("*No Seg. U-Net*" vs "*U-Net*" networks and "*No Seg. GAN*" vs "*GAN*" networks). The inclusion of the gradient restriction in the GAN topology also results in an improvement ("*Base GAN*" vs "*GAN*"), enabling the GAN to focus on improving the the information received by the 3D U-Net layers.

The reconstruction test performed using the BPF reconstruction showed a good correlation between the PET image reconstructed using the reference CT and the PET images obtained using the sCT. It can be seen in the difference maps (Figs. 8(d), 8(f), 10(d) and 10(f)) and in the profile plots (Figs. 9(a) and 11(a)) that the method presents differences with the reference in the lung area, probably due to the fact that the CT is taken instantly and the PET image is averaged over multiple respiration cylces. This is also reported in other studies [6, 7, 29]. Also it can be seen in Fig. 8(1) that the network fails to generate prosthesis structures present in the reference CT (Fig. 5(d)). This is observed as a negative error around the hip in Figs. 8(f), 8(d), 8(l) and 8(j). This is probably due to the lack of samples containing prosthesis in the train dataset. As stated before, this could be mitigated using MLAA techniques.

The metric scores of our methods are comparable to the method proposed by Dong [6]. Their method was trained using a dataset composed of NAC PET and CT co-registered samples from a ToF enabled PET-CT scanner. Their unsupervised method achieves a MAE of 108 ± 19 HU in the reconstruction of a sCT image. Nevertheless our models are more compact than their proposed cycle-GAN architecture, requiring less parameters. In the case of our GAN model, we control the adversarial gradient resulting in a more stable training procedure. Also our work is tested on multiple PET scanners. Moreover, most of the tested scanners do not posses ToF capabilities, for which other methods such as MLAA, result in lower quality whole body attenuation maps. Finally, we consider that it is important to posses a common dataset to enable direct comparison of different sCT generation

methods, in order to asses the required complexity of the models. For this reason the code and dataset used in this work is released ¹.

5 Conclusions

We presented two deep learning approaches to the task of attenuation map generation from uncorrected PET image data. The methods perform with image quality comparable to other methods and the attenuation sinograms difference is low, showing potential for PET image correction. The GAN method was able generate visually appealing images with high correlation to a real patient anatomy, however the 3D U-Net achieved better scores in the image quality and PET image correction metrics.

Competing interests

The authors have declared that no competing interests exist.

Funding

This work was supported by the Universidad Tecnológica Nacional, the Université de Technologie de Troyes, the Comisión Nacional de Energía Atómica and the National Scientific and Technical Research Council (CONICET).

Authors' contribution

RRC designed and implemented the network topologies, conducted the experimentation, analyzed the results and wrote the manuscript; CV and DM helped with conceiving the idea; DM designed the experiments; CV, DM and TG revised the results and manuscript. All authors read and approved the final manuscript.

References

- D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating ct image from mri data using 3d fully convolutional networks," in *Deep Learning and Data Labeling for Medical Applications*, pp. 170–178, Springer, 2016.
- [2] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "Medgan: Medical image translation using gans," *Computerized Medical Imaging and Graphics*, p. 101684, 2019.
- [3] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 14–23, Springer, 2017.
- [4] T. Wang, Y. Lei, Y. Fu, W. J. Curran, T. Liu, J. A. Nye, and X. Yang, "Machine learning in quantitative pet: A review of attenuation correction and low-count image reconstruction methods," *Physica Medica*, vol. 76, pp. 294–306, 2020.

- [5] F. Liu, H. Jang, R. Kijowski, G. Zhao, T. Bradshaw, and A. B. McMillan, "A deep learning approach for 18 f-fdg pet attenuation correction," *EJNMMI physics*, vol. 5, no. 1, p. 24, 2018.
- [6] X. Dong, Y. Lei, T. Wang, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, "Deep learningbased attenuation correction in the absence of structural information for whole-body pet imaging," *Physics in Medicine & Biology*, 2019.
- [7] K. Armanious, T. Hepp, T. Küstner, H. Dittmann, K. Nikolaou, C. La Fougère, B. Yang, and S. Gatidis, "Independent attenuation correction of whole body [18 f] fdg-pet using a deep learning approach with generative adversarial networks," *EJNMMI research*, vol. 10, no. 1, pp. 1–9, 2020.
- [8] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms," *IEEE transactions on medical imaging*, vol. 18, no. 5, pp. 393–403, 1999.
- [9] L. Shi, J. A. Onofrey, E. M. Revilla, T. Toyonaga, D. Menard, J. Ankrah, R. E. Carson, C. Liu, and Y. Lu, "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*, pp. 723–731, Springer, 2019.
- [10] J. Hamill and V. Panin, "Tof-mlaa for attenuation correction in thoracic pet/ct," in 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC), pp. 4040–4047, IEEE, 2012.
- [11] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [12] B. Tatiana, M. De Ornelas Couto, and I. B. Mihaylov, "Head-and-neck squamous cell carcinoma patients with ct taken during pre-treatment, mid-treatment, and posttreatment dataset.," 2018.
- [13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV)*, 2016 Fourth International Conference on, pp. 565–571, IEEE, 2016.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026– 1034, 2015.
- [15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the* 34th International Conference on Machine Learning (D. Precup and Y. W. Teh, eds.), vol. 70 of *Proceedings* of Machine Learning Research, (International Convention Centre, Sydney, Australia), pp. 214–223, PMLR, 06–11 Aug 2017.
- [16] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein

¹https://github.com/RawthiL/PET_DeepAttCorrection

gans," in Advances in neural information processing systems, pp. 5767–5777, 2017.

- [17] P. Bandi, N. Zsoter, L. Seres, Z. Toth, and L. Papp, "Automated patient couch removal algorithm on ct images," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 7783–7786, IEEE, 2011.
- [18] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, *et al.*, "Data for nsclc radiogenomics collection," *The Cancer Imaging Archive*, 2017.
- [19] M. L. Zuley, R. Jarosz, S. Kirk, Y. Lee, R. Colen, K. Garcia, and N. Aredes, "Radiology data from the cancer genome atlas head-neck squamous cell carcinoma [tcga-hnsc] collection," *Cancer Imaging Archive. doi*, 2016.
- [20] B. Albertina, M. Watson, C. Holback, R. Jarosz, S. Kirk, Y. Lee, and J. Lemmerman, "Radiology data from the cancer genome atlas lung adenocarcinoma [tcga-luad] collection," *Cancer Imaging Arch*, 2016.
- [21] S. Kirk, Y. Lee, C. Roche, *et al.*, "Radiology data from the cancer genome atlas thyroid cancer [tcgathca] collection," *Cancer Imaging Archive. doi*, 2016.
- [22] A. Meldo, L. Utkin, A. Lukashin, V. Muliukha, and V. Zaborovsky, "Database acquisition for the lung cancer computer aided diagnostic systems," in 2019 25th Conference of Open Innovations Association (FRUCT), pp. 220–227, IEEE, 2019.
- [23] A. B. Menegotto, C. D. L. Becker, and S. C. Cazella, "Computer-aided hepatocarcinoma diagnosis using multimodal deep learning," in *International Symposium on Ambient Intelligence*, pp. 3–10, Springer, 2019.
- [24] N. C. I. C. P. T. A. C. C. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), "Radiology data from the clinical proteomic tumor analysis consortium uterine corpus endometrial carcinoma (cptac-ucec) collection," 2019.
- [25] N. C. I. C. P. T. A. C. C. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CP-TAC), "Radiology data from the clinical proteomic tumor analysis consortium lung squamous cell carcinoma [cptac-lscc] collection," 2018.

- [26] J.-Y. Zhu, Z. Zhang, C. Zhang, J. Wu, A. Torralba, J. Tenenbaum, and B. Freeman, "Visual object networks: image generation with disentangled 3d representations," in *Advances in Neural Information Processing Systems*, pp. 118–129, 2018.
- [27] T. Nguyen-Phuoc, C. Li, L. Theis, C. Richardt, and Y.-L. Yang, "Hologan: Unsupervised learning of 3d representations from natural images," in *Proceedings* of the IEEE International Conference on Computer Vision, pp. 7588–7597, 2019.
- [28] L. Brusaferri, A. Bousse, N. Efthimiou, E. Emond, D. Atkinson, S. Ourselin, B. F. Hutton, S. Arridge, and K. Thielemans, "Potential benefits of incorporating energy information when estimating attenuation from pet data," in 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–4, IEEE, 2017.
- [29] X. Dong, T. Wang, Y. Lei, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, "Synthetic ct generation from non-attenuation corrected pet images for whole-body pet imaging," *Physics in Medicine & Biology*, vol. 64, no. 21, p. 215016, 2019.

Citation: R. R. Colmeiro, C. Verrastro, D. Minsky and T. Grosges. *Towards a Whole Body* [¹⁸*F*] *FDG Positron Emission Tomography Attenuation Correction Map Synthesizing using Deep Neural Networks*. Journal of Computer Science & Technology, vol. 21, no. 1, pp. 29-41, 2021. **DOI:** 10.24215/16666038.21.e04 **Received:** December 14, 2020 **Accepted:** March

26, 2021.

Copyright: This article is distributed under the terms of the Creative Commons License CC-BY-NC.