Cross-Modality Synthesis from CT to PET using FCN and GAN Networks for Improved Automated Lesion Detection

Avi Ben-Cohen[a,*], Eyal Klang[b], Stephen P. Raskin[b], Shelly Soffer[b], Simona Ben-Haim[c,d], Eli Konen[b], Michal Marianne Amitai[b,1], Hayit Greenspan[a,1]

[a]*Tel Aviv University, Faculty of Engineering, Department of Biomedical Engineering, Medical Image Processing Laboratory, Tel Aviv 69978, Israel*
[b]*Sheba Medical Center, Diagnostic Imaging Department, Abdominal Imaging Unit, affiliated to Sackler school of medicine Tel Aviv University, Tel Hashomer 52621, Israel*
[c]*Sheba Medical Center, Department of Nuclear Medicine, Tel Hashomer 52621, Israel*
[d]*Institute of Nuclear Medicine, University College London Hospital, London NW1 2BU, UK*

Abstract

In this work we present a novel system for generation of virtual PET images using CT scans. We combine a fully convolutional network (FCN) with a conditional generative adversarial network (GAN) to generate simulated PET data from given input CT data. The synthesized PET can be used for false-positive reduction in lesion detection solutions. Clinically, such solutions may enable lesion detection and drug treatment evaluation in a CT-only environment, thus reducing the need for the more expensive and radioactive PET/CT scan. Our dataset includes 60 PET/CT scans from Sheba Medical center. We used 23 scans for training and 37 for testing. Different schemes to achieve the synthesized output were qualitatively compared. Quantitative evaluation was conducted using an existing lesion detection software, combining the synthesized PET as a false positive reduction layer for the detection of malignant lesions in the liver. Current results look promising showing a 28% reduction in the average false positive per case from 2.9 to 2.1. The suggested solution is comprehensive and can be expanded to additional body organs, and different modalities.

Keywords: Deep learning, CT, PET, GAN, image synthesis, liver lesion

1. Introduction

The combination of positron emission tomography (PET) and computerized tomography (CT) scanners have become a standard component of diagnosis and staging in oncology [23, 13]. An increased accumulation of Fluoro-D-glucose (FDG) in PET relative to normal tissue is a useful marker for many cancers and can help in detection and localization of malignant lesions [13]. Additionally, PET/CT imaging is becoming an important evaluation tool for new drug therapies [25]. An example of an axial slice taken from a CT scan and its corresponding PET slice is shown in Figure 1. It can be seen that the PET image resolution is lower than the CT image resolution showing less anatomical details. However, there is a malignant liver lesion that is less visible in the CT image and can be easily detected in the PET image as a large dark blob.

Although PET imaging has many advantages and its use is steadily increasing, it has a few disadvantages. PET/CT entails added radiation exposure in comparison to CT-only scans. Moreover, PET/CT is relatively expensive compared to CT. Hence, it is still not offered in the large proportion of medical centers in the world. The clinical importance of PET in the management of cancer patients and on the other hand the difficulty in providing PET imaging as part of standard imaging, raises a potential need for an alternative, less expensive, fast, and easy to use PET-like imaging.

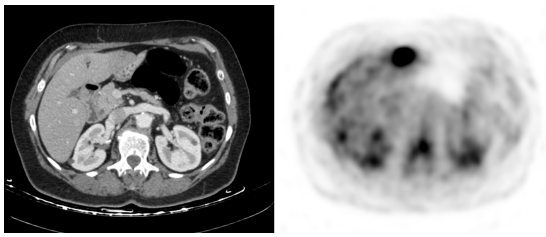

Figure 1: An axial CT slice (left) with its corresponding PET slice (right). Dark regions in the PET image indicate high FDG uptake.

Several works had recently explored cross-modality synthesis using deep learning methods [19, 10, 26]. In these works, different deep learning based methods and architectures were explored to learn an end-to-end nonlinear mapping from magnetic resonance images to CT images. For the case of unpaired data, a CycleGAN model was used to synthesize brain CT images from brain MR by Wolterink et al. [24]. Chartsias et al. [6] demonstrated a similar concept for synthesizing cardiac MR images from CT images. In the case of PET/CT pairs, the PET

*Corresponding author.
Email address: avibenc@mail.tau.ac.il (Avi Ben-Cohen)
[1]These authors contributed equally.

study can be used to highlight malignant lesions and improve the detection compared to the use of CT data alone. Bi et al. [5] used a multi-channel generative adversarial network that synthesizes PET images from CT images with manually annotated lung tumors. Their model learns the integration from both CT and a given annotated label, to synthesize the high uptake and the anatomical background. They have demonstrated using the synthesized PET images a comparable detection performance to that achieved using the original PET data. We note that manual labeling of the tumors is needed in this work.

In the current work our objective is to use information from CT data to estimate PET-like images with an emphasis on malignant lesions in the liver. The suggested system is fully automated, with no manual labeling needed. Similar to the radiologists, who have an easier time identifying malignant liver lesions in a PET/CT scan (vs only a CT scan) we want to make use of the estimated PET-like images to improve the detection of malignant lesions using an automated lesion detection software.

The proposed system is based on a fully convolutional network (FCN) and a conditional GAN (cGAN). The contributions of this work include: 1) We present a novel method to synthesize PET images from CT images, focused on malignant lesions with no manually labeled data; 2) The synthesized PET is shown to improve an existing automatic lesion detection software; 3) Reconstruction measures are presented for comparison between different methods.

This work is an extension to earlier work [4], in which we used a pyramid based image blending step to combine the advantages of an FCN and a cGAN network. In the current work we present a novel system architecture that obviates the need for an image blending step, thus providing savings in time and reducing the need for manually defining a threshold for the blending mask, while improving the system performance (as will be demonstrated in section 3). In addition, the dataset was substantially extended.

To achieve the virtual PET we use advanced deep learning techniques with both fully convolutional networks and conditional adversarial networks as described in the following subsections.

1.1. Fully Convolutional Networks

In recent years, deep learning has become a dominant research topic in numerous fields. Specifically, Convolutional Neural Networks (CNN) have been used for many challenges in computer vision. CNN obtained outstanding performance on different tasks, such as visual object recognition, image classification, hand-written character recognition and more. Deep CNNs introduced by LeCun et al. [17], is a supervised learning model formed by multi-layer neural networks. CNNs are fully data-driven and can retrieve hierarchical features automatically by building high-level features from low-level ones, thus obviating the need to manually customize hand-crafted features. Previous works have shown the benefit of using a fully convolutional architecture for liver lesion detection and segmentation applications [2, 8]. FCNs can take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. Unlike patch based methods, the loss function using this architecture is computed over the entire image. The network processes entire images instead of patches, which removes the need to select representative patches, eliminates redundant calculations where patches overlap, and therefore scales up more efficiently with image resolution. Moreover, there is a fusion of different scales by adding links that combine the final prediction layer with lower layers with finer strides.

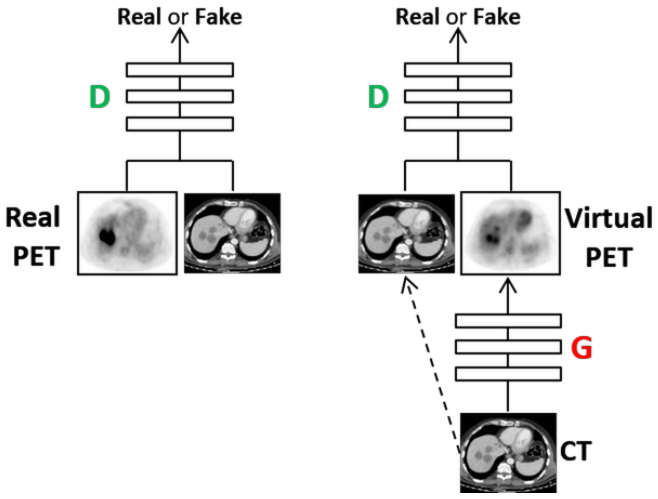

Figure 2: Training a cGAN to predict PET images from CT images. The discriminator, D, learns to classify between real and synthesized pairs. The generator, G, learns to fool the discriminator.

1.2. Conditional Adversarial Networks

More recent works show the use of Generative Adversarial Networks (GANs) for image to image translation [12]. GANs are generative models that learn a mapping from random noise vector z to output image y [9]. In contrast, conditional GANs (cGANs) learn a mapping from observed image x and random noise vector z, to y. The generator G is trained to produce outputs that cannot be distinguished from real images by an adversarially trained discriminator, D, which is trained to detect real vs fake images. Figure 2 shows a diagram of this procedure.

In this study we use FCN and cGAN to generate PET-like images from CT volumes. The strengths of both methods are used to create realistic looking virtual PET images. We focus our attention to hepatic malignant lesions. The method is presented in Section 2. Experiments and results are described in Section 3. The experiments include comparison of various algorithmic solutions to the task along with an evaluation of the benefit of our method to an existing automatic liver lesion detection software. We conclude this paper with a discussion in Section 4.

2. Methods

Our framework includes two main modules: a training module which includes data preparation, and a testing module which

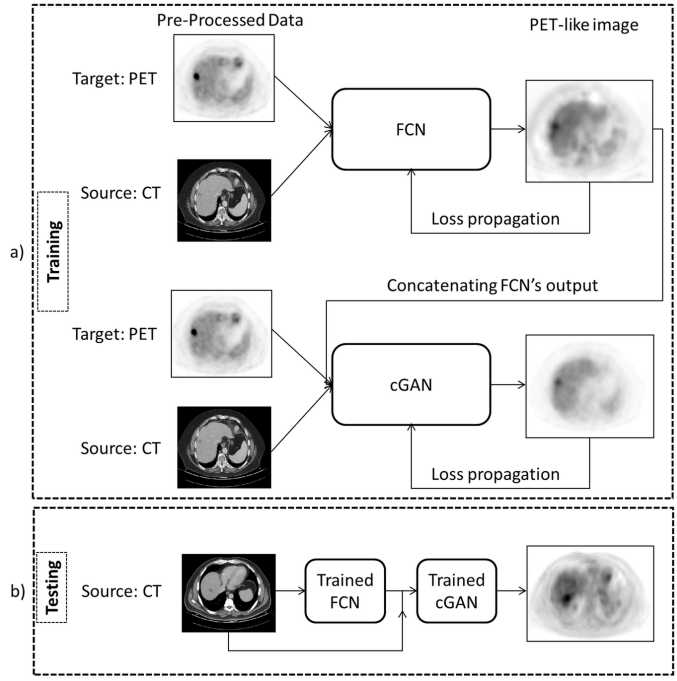

Figure 3: The proposed virtual PET system.

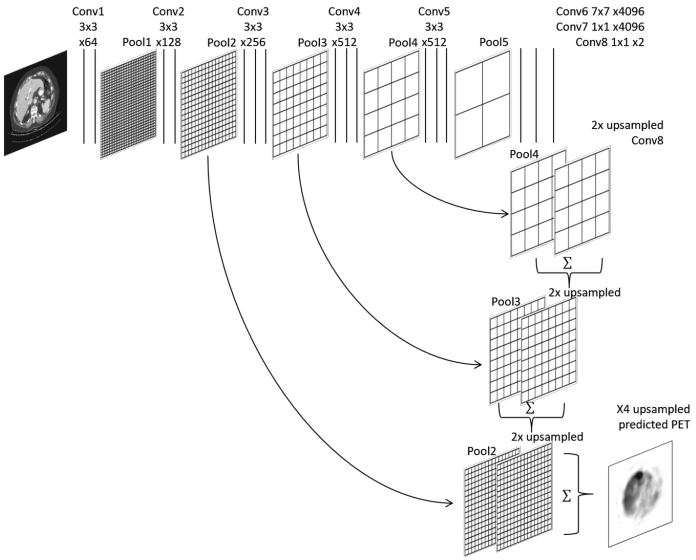

Figure 4: FCN-4s architecture. Each convolution layer is illustrated by a straight line with the receptive field size and number of channels denoted above. The ReLU activation function and drop-out are not shown for brevity.

accepts CT images as input and predicts synthesized PET images. We use an FCN to generate initial PET-like images given the input CT images. We next use a cGAN to improve and refine the FCN output. Figure 3 shows a diagram of our general framework. Each module will be described in depth in the following subsections.

2.1. Training Data Preparation

The training input for the FCN includes two image types: a source CT image and a target PET image. A similar size is needed for the two images and in most cases the PET resolution is much lower than the CT. Hence, the first step in preparing the data for training was to align the PET scans with the CT scans using the given offset, pixel-spacing and slice-thickness of both scans using linear interpolation. Secondly, we wanted to limit our PET values to a constrained range of values. The standardized uptake value (SUV) is commonly used as a relative measure of FDG uptake [11] as in equation 1:

$$SUV = \frac{r}{a'/w} \tag{1}$$

where r is the radioactivity concentration [kBq/ml] measured by the PET scanner within a region of interest (ROI), a' is the decay-corrected amount of injected radiolabeled FDG [kBq], and w is the weight of the patient [g], which is used as a surrogate for a distribution volume of tracer.

CT and PET studies include a large value range. To assist the network to learn the translation between these modalities, we found experimentally that some constraints were helpful: We used contrast adjustment, by clipping extreme values and scaling, to adjust the PET images into the SUV range of 0 to 20. This range includes most of the interesting SUV values of malignant lesions. Similarly, CT image values were adjusted to be within -160 HU to 240 HU (Hounsfield Units); the standard HU windowing used by the radiologists when evaluating the liver parenchyma.

2.2. Fully Convolutional Network Architecture

In the following we describe the FCN used for both training and testing as in Figure 3a and 3b. The FCN network architecture uses the VGG 16- layer net [22]. We convert all fully connected layers to convolutions and remove the classification layer. We append a 1x1 convolution with channel dimension to generate the PET-like images. Upsampling is performed in-network for end-to-end learning by backpropagation from the pixelwise L_2 loss. The FCN-4s net was used as our network, which learned to combine coarse, high layer information with fine, low layer information as described in [21] with an additional skip connection by linking the Pool2 layer in a similar way to the linking of the Pool3 and Pool4 layers in Figure 4.

2.3. cGAN Architecture

The output from the FCN was found experimentally to have a good response in regions with high SUV but to be less accurate and blurry in regions with low contrast. Hence, cGAN was used to refine the FCN's output. The input to the cGAN included two channels, one with the CT image and the second with the corresponding FCN output. The training concept follows the same trend of the training process of the FCN optimizing the network's generator based on the pixelwise L_2 loss as well as the cross entropy classification error for the discriminator, as will be further elaborated in section 2.4. We adapt a similar cGAN architecture as in [12] with a few modifications. In the original cGAN the objective can be expressed as:

Table 1: The "U-Net" based generator used in the proposed cGAN architecture.

U-Net encoder			U-Net decoder		
Layer	**Details**	**Size**	**Layer**	**Details**	**Size**
input	CT image; FCN's output	512x512 x2	upsampling1	2x2 upsample of conv5_2 concatenate with conv4_2	64x64 x768
conv1_1	3x3x32; dilaton rate 3; LeakyReLU	512x512 x32	conv6_1	3x3x256; LeakyReLU	64x64 x256
conv1_2	3x3x32; dilaton rate 3; LeakyReLU	512x512 x32	conv6_2	3x3x256; LeakyReLU	64x64 x256
pool1	2x2 max pool; stride 2	256x256 x32	upsampling2	2x2 upsample of conv6_2 concatenate with conv3_2	128x128 x384
conv2_1	3x3x64; dilaton rate 2; LeakyReLU	256x256 x64	conv7_1	3x3x128; LeakyReLU	128x128 x128
conv2_2	3x3x64; dilaton rate 2; LeakyReLU	256x256 x64	conv7_2	3x3x128; LeakyReLU	128x128 x128
pool2	2x2 max pool; stride 2	128x128 x64	upsampling3	2x2 upsample of conv7_2 concatenate with conv2_2	256x256 x192
conv3_1	3x3x128; LeakyReLU	128x128 x128	conv8_1	3x3x64; dilaton rate 2; LeakyReLU	256x256 x64
conv3_2	3x3x128; LeakyReLU	128x128 x128	conv8_2	3x3x64; dilaton rate 2; LeakyReLU	256x256 x64
pool3	2x2 max pool; stride 2	64x64 x128	upsampling4	2x2 upsample of conv8_2 concatenate with conv1_2	512x512 x96
conv4_1	3x3x256; LeakyReLU	64x64 x256	conv9_1	3x3x32; dilaton rate 3; LeakyReLU	512x512 x32
conv4_2	3x3x256; LeakyReLU	64x64 x256	conv9_2	3x3x32; dilaton rate 3; LeakyReLU	512x512 x32
pool4	2x2 max pool; stride 2	32x32 x256	conv10	1x1x1	512x512 x1
conv5_1	3x3x512; LeakyReLU	32x32 x512			
conv5_2	3x3x512; LeakyReLU	32x32 x512			

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{ct,pet}[\log D(ct,pet)] + \mathbb{E}_{ct,z}[\log(1 - D(ct, G(ct,z))] \quad (2)$$

where G tries to minimize this objective against an adversarial D that tries to maximize it, ct is the CT input slice, pet is the corresponding PET slice, and z is random noise. In our objective we embed the FCN's output (fcn) as in:

$$\begin{aligned}\mathcal{L}_{Modified-cGAN}(G,D) =& \mathbb{E}_{fcn,ct,pet}[\log D(fcn,ct,pet)] + \\ & \mathbb{E}_{fcn,ct,z}[\log(1 - D(fcn,ct,G(fcn,ct,z))]\end{aligned} \quad (3)$$

We tried both $L1$ and $L2$ distance measures for the generator G. No noticeable difference was observed using the different distance measures and we chose to use the $L2$ in our experiments. The final optimization process:

$$G^* = arg\min_G \max_D L_{Modified-cGAN}(G,D) + \lambda \mathbb{E}_{fcn,ct,z}\|t - G(l,c)\|_2 \quad (4)$$

where G^* is the optimal setting and λ balances the contribution of the two terms.

Table 1 and Table 2 show the different components of the network for the "U-Net" [20] based generator (encoder and decoder) and the discriminator respectively.

2.4. Loss Weights

Malignant lesions are usually observed with high SUV values (> 2.5) in PET scans [16]. Hence, we used the SUV value in each pixel as a weight for the pixel-wise loss function as in equation (5) where N is the number of samples. By this we allow the network to pay more attention to high SUV even though most pixels include lower values.

Table 2: The discriminator used in the proposed cGAN architecture.

Discriminator		
Layer	Details	Size
input	Includes: real/fake PET image; CT image;FCN's output	512x512x3
conv1	3x3x32; stride 2; LeakyReLU	256x256x32
conv2	3x3x64; stride 2; LeakyReLU	128x128x64
conv3	3x3x128; stride 2; LeakyReLU	64x64x128
conv4	3x3x256; stride 1; LeakyReLU	64x64x256
dense	2 classes (fake/real);softmax	2

$$L = \frac{1}{N}\sum_{i=1}^{N} I_{PET}(i)(Syn_{PET}(i) - I_{PET}(i))^2 \quad (5)$$

While this approach helped the FCN to learn the malignant tumor appearance and provide a better response in the synthesized PET images, it did not help when training the cGAN. Hence, we modified the loss function by separating it into high SUVs (> 2.5) and low SUVs (≤ 2.5) as in equation (6). This way the cGAN training was able to achieve better response in regions with high SUV while not hurting other regions.

$$L_{cGAN} = L_{lowSUV} + L_{highSUV} \quad (6)$$

3. Experiments and Results

To evaluate the performance of the system, we conducted several sets of experiments. A development set was used for training and validation. Final results are presented for an independent testing set. We demonstrate the applicativity of the system to augment performance of a liver lesion detection system.

3.1. Dataset

An institutional review board (IRB) approval was granted for this retrospective study and informed consent was waived by the IRB committee. The data used in this work includes PET/CT scans (pairs of contrast enhanced portal phase CT scans with their corresponding PET scans) from the Sheba Medical Center, obtained from 2014 to 2015. The dataset contains 60 CT (with 0.97 mm pixel spacing and 4 mm slice thickness) and PET (with 3 mm pixel spacing and 4 mm slice thickness) pairs (from 60 different patients) which we constrained to slices in the region of the liver for our study. Not all PET/CT scans in our dataset included liver lesions. The training set included 23 PET/CT pairs (6 with malignant liver lesions) and the testing was performed on 37 pairs (9 with malignant liver lesion).

3.2. Experimental Setting

The networks were implemented and trained using Keras framework [7] on a PC with a single NVIDIA 1080 Ti GPU.

The following hyper-parameters were used for all networks: learning rate of 0.00001 with a batch size of 4. Adam optimizer [15] was used with $\beta = 0.5$. For the cGAN, we used $\lambda = 20$ in the optimization process presented in equation 4. To assist the network model to be more robust to to variability in location and scaling, online data transformations were performed with uniform sampling of scale [-0.9,1.1] and translations [-25,25] in each epoch.

3.3. Reconstruction Evaluation

To quantitatively evaluate our method performance in means of reconstruction we used the mean absolute error:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Syn_{PET}(i) - I_{PET}(i)| \quad (7)$$

where i iterates over aligned voxels in the real and synthesized CT images.

In addition, the peak-signal-to-noise-ratio (PSNR) as in [19, 24] was used:

$$PSNR = 20\log_{10}\frac{20}{MSE} \quad (8)$$

where MSE is the mean-squared error, i.e. $\frac{1}{N}\sum_{i=1}^{N}(Syn_{PET}(i) - I_{PET})^2$.

Our goal is to get very good reconstruction in the lesion areas, while keeping high reconstruction quality within the entire scan. High SUV values within a PET scan often serve as an indicator for the malignant lesions. We therefore measure the reconstructive error for the high SUV values (larger than 2.5) and the low SUV values, as two seperate sets. The average of these measures is computed as the final score.

Table 3 shows quantitative comparison across several possible reconstruction schemes. Our proposed method was compared to two different fully convolutional networks: A U-net based model as in 1 and the FCN-4s as in 4. We used the loss as in equation 5 since both of the fully convolutional based architectures provided better results compared to the loss proposed in equation 6, which was tailored for the cGAN architecture. We have also compared our proposed method using the loss as in equation 6 to the loss as in equation 5. In addition, we have compared the results to the cGAN as in figure 2 (without using the output of the FCN-4S) and to another method which used pyramid based image blending to combine the FCN-4s output with the cGAN [4]. Our proposed method showed superiority over the other methods with an average MAE of 0.72 and 0.79, and PSNR of 30.22 and 30.4 using the loss as in equation 6 and equation 5 respectively. When using the loss as in equation 5 instead of equation 6 our method achieved a better average PSNR, however, we preferred to use the latter since it achieved better reconstruction measurements for the high SUVs.

Qualitative results are shown in Figure 5. It compares the method's virtual PET images with the original PET study images. The virtual PET provided a very similar response to the real PET in the presented cases. The left column includes one malignant lesion, the second column includes three liver metastases (malignant lesions), and the right column includes two cysts (benign lesions). In Figure 6 we have employed the same sample cases as in Figure 5 and have compared our proposed method to the image blending based method [4]. In the left column we can see that the malignant lesion has been recognized using both methods as a dark blob which seems slightly darker and larger using our method. In the second column the three malignant lesions were recognized using both methods, however, the image blending based method included several blobs that are false-positives marked in red. In the right column there are two cysts (benign lesions), and both methods did not have high response as expected, however, our proposed method seems to have a better result for the surrounding tissues such as the left kidney marked in green.

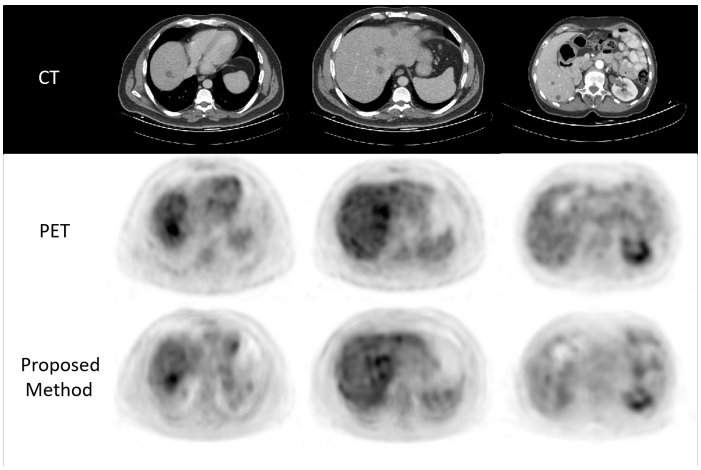

Figure 5: Sample results of the predicted PET using our method compared to the real PET with the corresponding CT images.

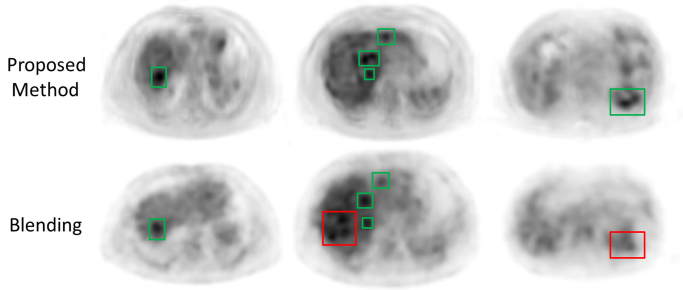

Figure 6: Sample results of the predicted PET using our method compared to the image blending based method [4]. In green - correctly synthesized high SUV regions; In red - false synthesized high SUV regions.

Table 3: Average reconstruction performance for low and high SUV regions using different methods. In bold - the highest score in each column.

Method	High SUV		Low SUV		Average Score	
	MAE	PSNR	MAE	PSNR	MAE	PSNR
Proposed Eq. 6	**1.33**	22.40	0.11	38.04	**0.72**	30.22
Proposed Eq. 5	1.48	21.7	**0.09**	**39.1**	0.79	**30.4**
blending	1.50	21.40	0.10	39.00	0.80	30.20
GAN	1.70	20.62	0.10	39.06	0.90	29.84
FCN-4s	**1.33**	**22.50**	0.16	37.60	0.74	30.05
UNET	1.52	21.57	0.12	38.56	0.82	30.07

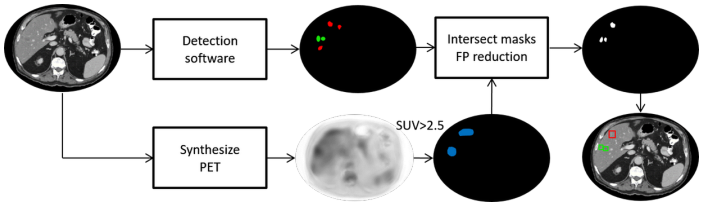

Figure 7: Combining our proposed method to synthesize PET with a lesion detection software: The output of the detection software includes true positives (in green) along with false positives (in red). Thresholding over the synthesized PET image to extract high-response regions (in blue) can help reduce the false positives by intersecting the detection mask with the thresholding mask.

3.4. *Liver Lesion Detection using the Virtual-PET*

In the following experiment, we use the synthesized (Virtual) PET images as an additional false-positive reduction layer for an existing lesion detection software. Since high SUV values can be indicative of malignant lesions, thresholding the PET using a high SUV threshold ($th = 2.5$) can reduce non-relevant regions for lesion detection. We use a particular lesion detection system, that was developed in our group [3], which combines global context via an FCN, along with local patch level analysis using superpixel sparse based classification. This framework is made up of two main modules. The first module is an FCN having three slices are as input: the target slice in the center and two adjacent slices above and below. Using an FCN - based analysis, this module outputs a lesion probability map. Based on the high-probability candidate lesion regions from the first module, the second module follows with localized patch level analysis using superpixel sparse based classification. This module's objective is to classify each localized superpixel as a lesion or not. Thus it provides a fine-tuning step, with the objective of increasing sensitivity to lesions while removing false positives (FPs).

We suggest the following scheme, for improving the lesion detection software: Given a CT scan, the detection software outputs lesion candidates as a binary mask which may include false detections. By thresholding the synthesized PET scan and finding the intersection of both the detection software's candidates mask and the PET thresholding result some of the false detections can be removed. Figure 7 illustrates this process. This is a rather naïve approach that shows the clinical relevance of using the virtual PET to improve existing software.

An additional test set of 14 CT scans including 55 lesions was used for the following experiments. Two evaluation measurements were computed, the true positive rate (TPR) and false positive rate (FPR) for each case as follows:

- *TPR*- Number of correctly detected lesions divided by the total number of lesions.
- *FPR*- Number of false positives per scan.

Table 4 shows the performance with and without the proposed false-positive reduction layer (note that the results were constrained to the manually annotated liver). Using our method the average FPR decreased from 2.9 to 2.1 (improvement of 28% with P-value<0.05) with a similar TPR.

Table 4: Detection measurements with and without SUV thresholding on the synthesized PET. In bold - the results obtained using the proposed method.

Method	TPR[%]	FPR
Detection soft.	94.6	2.9
Detection soft+ proposed	**94.6**	**2.1**
Detection soft+ blending	90.9	2.2
Detection soft+ FCN-4s	90.9	2.2

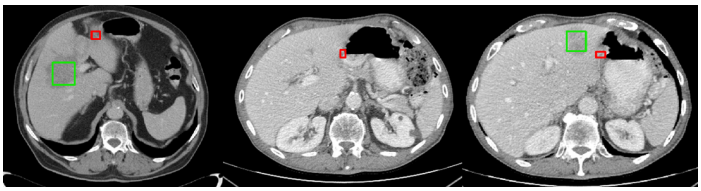

Figure 8: Sample results using the existing detection software. In green - correctly detected lesions. In red - false positives that were removed by combining the proposed method.

Figure 8 shows examples of cases with false positives that were removed (in red) by combining the proposed method with the existing detection software.

One key parameter of the detection system is the probability threshold, *th*, which defines the set of candidate regions extracted from the FCN probability maps. We compare the free-response receiver operating characteristic (FROC) presented in [3] with the FROC achieved using the combination of the synthesized PET with the current system (Figure 9). It can be seen that our system was able to reduce the amount of false positives in each tested *th* while preserving the TPR. Unlike the classic FROC curve where the TPR increases for decreasing *th*, here there was a moderate decrease for *th* below 0.95. The TPR can decrease when two candidates merge into one candidate even though there are two lesions. Note that We used $th = 0.95$ in our experiments (table 4 and figure 8) as in the original system.

As a final experiment, we tested a fully automatic framework with an automatic liver segmentation scheme, as in [3]. Thus, we use an automatic liver segmentation instead of the manually circumscribed liver. Using our method the average

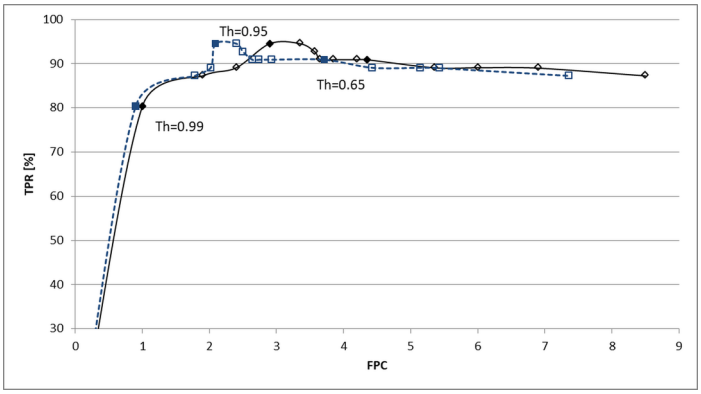

Figure 9: FROC curve of lesion detection using FCN with sparsity based FP reduction (solid black line) and with the addition of synthesized PET FP reduction (dashed blue line).

FPR decreased from 3.0 to 2.3 (improvement of 23% with P-value<0.05) with a slight decrease of the TPR from 90.9% to 89.1% (not significant).

4. Discussion

A novel system for PET synthesis using only CT scans has been presented. The presented system includes an FCN model and a cGAN model that refines the synthesized output extracted from the FCN. This framework provides a realistic PET estimation with special attention to malignant lesions using a custom loss function for each model.

Table 3 shows the reconstruction performance in terms of MAE and PSNR for high SUV regions, low SUV regions, and an average of both which is an estimate of the balance we want to have between these regions. The high SUV regions are important since they usually signify malignant lesions inside the liver, but it is important to reconstruct the low SUV regions as well, since we want to have a good contrast between malignant and non-malignant tissues. The FCN-4s achieved the best PSNR and MAE for the high SUV regions with our proposed method achieving very close performance measures. However, for the low SUV regions, the FCN-4s got inferior PSNR and MAE, while the cGAN and the image blending got the best results. Using the average scores our proposed method seems to have the best balance between the high and low SUV regions. Another method that was tested is the U-Net [20]. From Table 3 it can be seen that the U-Net did not provide better results than the FCN-4s, hence we chose to use the FCN-4s as the initial network model.

In an additional experiment we wanted to see if the proposed method can be used to improve an existing automatic liver lesion detection software [3]. Our system was easily integrated into the lesion detection software. Using a pathological SUV threshold of 2.5 we achieved a decrease in false-positive from an average of 2.9 per case to 2.1 (28% improvement). This experiment shows the benefit of using the proposed system to improve a given liver lesion analysis software. Since this method was trained on a dataset that was not seen by the existing detection software it improved its results. However, no manual labeling was conducted in this experiment since our method uses only PET/CT pairs for training.

The experiments conducted in this study focused on the liver region. The liver is a common site for metastases in oncological patients, hence, more examples of malignancies can be extracted and can help the training process. More work should be done to make use of the proposed framework for other malignancies within different regions in the CT scan, preferably with no manually annotated labels, as presented in this work.

One possible application could be to use the virtual PET to improve lesion segmentation. However, the PET images are quite blurry and so is the virtual PET, making it hard to assess the segmentation process. Hence, we believe that detection approaches are more relevant for this method.

We used a rather naïve approach by thresholding the virtual PET to reduce the amount of false-positives per case in an existing lesion detection software. However, the proposed system can be easily integrated into the training process of different networks for different tasks such as detection (as shown here) and classification.

To conclude, our proposed framework with the FCN-cGAN combination and the custom loss function has shown promising results in terms of reconstruction measures as well as detection measures by integrating it with an existing lesion detection software. A major strength of this paper is that no manual labeling was used to train the system. As we well know, the task of manually labeling and annotating medical data is hard, and contributes to the usually small data sets that are used in todays medical imaging research. Each year millions of PET/CT studies are conducted worldwide, and utilizing the current method, the CT and PET pairing can be used as free labeled and annotated data, with potential for big data, approaching millions of studies. Future work entails obtaining a larger dataset with vast experiments using the entire CT and not just the liver region as well as integrating it into the training process of deep learning based detection and classification networks. The presented system can be used for many applications in which PET examination is needed such as evaluation of drug therapies and detection of malignant lesions.

Acknowledgment

This research was supported by the Israel Science Foundation (grant No. 1918/16).

Avi Ben-Cohen's scholarship was funded by the Buchmann Scholarships Fund.

Disclosure of conflict of interest

The authors have no relevant conflicts of interest to disclose.

References

[1] Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., and Ogden, J. M.,1984. Pyramid methods in image processing. RCA engineer, 29(6), pp. 33-41.

[2] Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., and Greenspan, H., 2016. Fully Convolutional Network for Liver Segmentation and Lesions Detection. In International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Springer International Publishing, pp. 77-85.

[3] Ben-Cohen, A., Klang, E., Kerpel, A., Konen, E., Amitai, M. M., and Greenspan, H., 2017. Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations. Neurocomputing.

[4] Ben-Cohen, A., Klang, E., Raskin, S. P., Amitai, M. M., and Greenspan, H.,2017. Virtual PET Images from CT Data Using Deep Convolutional Networks: Initial Results. In International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, pp. 49-57.

[5] Bi, L., Kim, J., Kumar, A., Feng, D., and Fulham, M., 2017. Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs). In Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Springer, Cham, pp. 43-51.

[6] Chartsias, A., Joyce, T., Dharmakumar, R., and Tsaftaris, S. A., 2017. Adversarial Image Synthesis for Unpaired Multi-modal Cardiac Data. In International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, pp. 3-13.

[7] Chollet, François et al.: Keras. `https://github.com/keras-team/keras`. GitHub, (2015).

[8] Christ, P. F., Ettlinger, F., Grn, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., ... and Rempfler, M., 2017. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. arXiv preprint arXiv:1702.05970.

[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., and Bengio, Y.,2014. Generative adversarial nets. In Advances in neural information processing systems, pp. 2672-2680.

[10] Han, X.,2017. MR based synthetic CT generation using a deep convolutional neural network method. Medical Physics, 44(4), pp. 1408-1419.

[11] Higashi, K., Clavo, A. C., and Wahl, R. L.,1993. Does FDG Uptake Measure the Proliferative Activity of Human Cancer Cells? In Vitro Comparison with DNA Flow Cytometry and Tritiated Thymidine Uptake. Journal of Nuclear Medicine, 34, 414-414.

[12] Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A.,2016. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004.

[13] Kelloff, G. J., Hoffman, J. M., Johnson, B., Scher, H. I., Siegel, B. A., Cheng, E. Y., and Shankar, L., 2005. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clinical Cancer Research, 11(8), 2785-2808.

[14] Kinehan, P. E., and Fletcher, J. W., 2010. PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. Semin Ultrasound CT MR, 31(6), 496-505.

[15] Kingma, D. P., and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[16] Kostakoglu, L., Agress Jr, H., and Goldsmith, S. J., 2003. Clinical role of FDG PET in evaluation of cancer patients. Radiographics, 23(2), 315-340.

[17] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.,1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[18] Metz, C. E., 2006. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. Journal of the American College of Radiology, 3(6), 413-422.

[19] Nie, D., Cao, X., Gao, Y., Wang, L., and Shen, D., 2016. Estimating CT image from MRI data using 3D fully convolutional networks. In International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. Springer International Publishing, pp. 170-178.

[20] Ronneberger, O., Fischer, P., and Brox, T.,2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer International Publishing, pp. 234-241.

[21] Shelhamer, E., Long, J., and Darrell, T., 2016. Fully convolutional networks for semantic segmentation. IEEE transactions on pattern analysis and machine intelligence.

[22] Simonyan, K., and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[23] Weber, W. A., Grosu, A. L., and Czernin, J., 2008. Technology Insight: advances in molecular imaging and an appraisal of PET/CT scanning. Nature Clinical Practice Oncology, 5(3), 160-170.

[24] Wolterink, J. M., Dinkla, A. M., Savenije, M. H., Seevinck, P. R., van den Berg, C. A., and IÅgum, I., 2017. Deep MR to CT synthesis using unpaired data. In International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, pp. 14-23.

[25] Weber, W. A., 2009. Assessing tumor response to therapy. Journal of nuclear medicine, 50(Suppl 1), 1S-10S.

[26] Xiang, L., Wang, Q., Nie, D., Qiao, Y., and Shen, D., 2017. Deep Embedding Convolutional Neural Network for Synthesizing CT Image from T1-Weighted MR Image. arXiv preprint arXiv:1709.02073.