서울대학교 대학원 융합전공 혁신의과학 전공

학부연구생인턴 보고서

연 구 기 간: 2022. 06. 20 ~ 2022. 08. 08

연 구 제 목: Brain PET image denoising

성 명: 박서연 (인 또는 서명) 박서연

지도교수 성 명: 이재성 (인 또는 서명)

(※ A4용지 10장 이내, 바탕체, 11포인트, 줄간격 160%. 아래 항목대로 기술하기 힘든 경우에는 해당 항목만을 기술함)

1. Abstract (영문의 경우 : 150단어 이내, 국문의 경우 : 750자 이내)

The detection and quantitative accuracy of PET imaging is limited by low spatial resolution and high noise. To improve PET image quality, eliminating noise is crucial. Recently, deep neural network-methods have made remarkable outcomes in denoising images. Our study aims to compare the performance of PET image denoising based on different convolutions and learning algorithms. We conducted experiments with ^{18}F-Florbetaben brain PET data of 47 patients. All of the methods performed effectively for image quality enhancement. While N2N gave comparable PSNR and SSIM to N2C, 3D convolution gave higher PSNR and SSIM than 2D convolution, resulting a denoised image with better details. The self-supervised method and 3D convolution approach for PET image denoising are useful to reduce the PET scan time.

2. Introductions (Research Background and Purpose)

Positron emission tomography (PET) is a functional imaging modality that uses compounds labeled with positron emitting radioisotopes as molecular tracers to image metabolic and biochemical processes in the body. PET is widely used for diagnosis, staging, therapy monitoring, and prognosis in oncology, cardiology, and neurology. However, low spatial resolution and signal-to-noise ratio limits the detection and quantitative accuracy of PET imaging. The noise in PET images is due to variety of physical degradation factors such as acollinearity, crystal penetration, and positron range. In addition, the finite number of detected photons during a given scan time adds statistical noise. To reduce radiation exposure to patients, the amount of radiotracer injection is limited, which gives us smaller number of counts and eventually higher noise level. Although the longer scan time of PET can increase the number of photon counts, the probability of image blurring caused by motion also increases.

To improve PET image quality, eliminating noise is indispensable and has become a crucial pre-processing step for low-dose PET imaging. Over the past decades, post-filtering approaches were proposed for PET image denoising, such as HYPR processing, non-local mean (NLM), wavelet, and adaptive diffusion filtering. Recently, deep neural networks (DNNs) have been applied to medical imaging, and demonstrated comparable or superior results to the traditional methods but at a faster speed.

PET image is a three-dimensional, volumetric medical data. For the application of deep learning in 3D medical images such as PET images, there are 2D-based approaches and 3D-based approaches. 2D approaches use slices of 3D images (transaxial, coronal, and sagittal) and extract features using 2D convolutions. Although large scale of 2D datasets are available for training, 2D convolutions have limitations in learning large 3D contexts. On the other hand, 3D approaches learn volumetric representations by using 3D convolutions, therefore large 3D contexts can be captured in any axis. However, data scarcity and computational cost problem emerges in 3D medical images.

When applying deep learning to image denoising techniques, learning algorithms should also be considered together. Early techniques used the fully supervised model, in other words Noise2Clean (N2C), which learned to map corrupted observations to clean signals. Although this algorithm may be simple, obtaining large number of clean training targets is difficult. Therefore, recent approaches in denoising focus on self-supervised learning such as Noise2Noise (N2N), which learns to restore images by only using corrupted images as input-target pairs.

In this study, we applied different deep learning methods for denoising PET images. We compared the denoising performance of 2D and 3D convolution methods, and also supervised and self-supervised methods to show the comparison between different learning algorithms.

3. Method

3.1 Dataset

We conducted experiments with ^{18}F-Florbetaben brain PET data of 47 patients (40 for training and 7 for testing). ^{18}F-Florbetaben is a radiotracer which is highly specific and sensitive for the β-amyloid neuritic plaque density.

The intake of β-amyloid deposits in ^{18}F-FBB is used for the diagnosis of Alzheimer's disease.

The list-mode PET data were retrospectively acquired with a Biograph mCT40 scanner (Siemens Healthcare, Knoxville, TN), 90min after the intravenous injection of ^{18}F-FBB (305.9MBq/kg) for 10 min in a single bed position. From the list-mode data, 20 data bins with 30 second duration and 600 second reference data were generated. The matrix and pixel sizes of the PET images were $200 \times 200 \times 109$ and $2.04 \times 2.04 \times 2.03\,mm^3$, respectively.

3.2 Network Architecture and training

In this study, two different U-Net models (2D U-Net, 3D U-Net) were each employed to train and test two different denoising methods (N2C, N2N): 2D N2C, 2D N2N, 3D N2C, 3D N2N.

U-Net is a CNN-based model widely used for tasks involving biomedical imaging such as image segmentation. The network consists of an encoder of contracting path to capture the context of an image, and a decoder with an expansive path that helps precise localization. For each layer of the network, 3×3 convolutions and ReLU are repeated twice, and 2×2 max pooling operation with stride 2 is used for downsampling. In the expanding path, feature map is upsampled by a factor of 2. The cropped feature map from the contracting path is then concatenated correspondingly, and 3×3 convolutions and ReLU functions are repeated twice. The 3D U-Net is proposed by replacing the 2D operations such as 2D convolution and 2D maxpooling with their 3D counterparts.

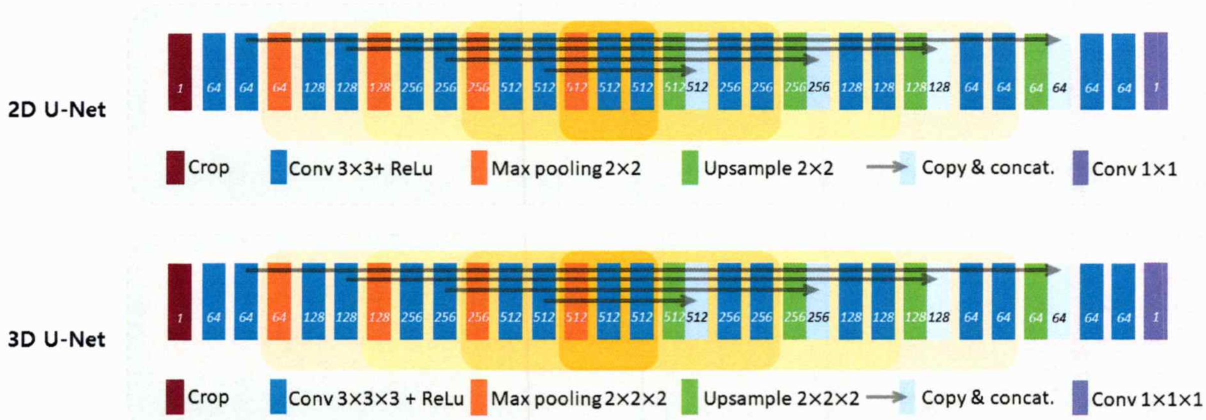

Figure 1. 2D/3D U-net structure. 3D U-net is implemented by replacing the 2D convolutions with 3d convolutions.

For training 2D U-Net, a total of 150 slices were randomly chosen from each patient, 50 slices each of 200×200 size transaxial slices, and 200×109 size coronal and sagittal slices. To apply maxpooling 4 times, each of the slices were then cropped into 192×192 and 192×96, respectively. For the evaluation, the slices from three different views were averaged. the 3D U-Net generated 50 random image patches of size $32 \times 32 \times 32$ from each patient to train the network. 30 sec images were used as input images and 600 sec images were used as target images in N2C. Independent 30 sec images were used as input and target images in N2N. For both methods, L2 norm between the output and target was used as the loss function to be minimized. The training samples were divided by the 99th percentile value of each image for normalization. To avoid learning from background, the percentage of pixel intensity over zero was calculated for each image and 50 percent was given as a threshold.

We implemented the training and testing of the network using PyTorch. The networks were trained for 60 epochs with a batch size of 16, and Adam Optimizer was used for optimization.

3.3 Image Analysis

We used peak signal-to-noise ratio (PSNR) and Structural Similarity Index (SSIM) for image quality assessment. 600 sec scan images were used as the reference image to calculate the PSNR and SSIM values.

PSNR is a common method to measure the fidelity that is independent of the dynamic range of images.

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (1)$$

In equation (1), the MAX value is the maximum value of the reference image, and MSE is the mean square error between the tested and reference image. Although PSNR is simple to calculate, it does not match well with the perceived visual quality.

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

SSIM is therefore used to take advantage of the human visual system, assessing image quality by extracting structural information and calculating the similarity. The system measures the similarity using three comparisons:

luminance, contrast, and structure. In equation (2), x and y are the images to be compared, and μ and σ refer to the mean and variance of the images. Constant C_1 and C_2 are included to avoid instability and are calculated as $C_1=(K_1L)^2$ and $C_2=(K_2L)^2$. L is the dynamic range of the pixel values, and $K_1=0.01$ and $K_2=0.03$ were used as small constants.

4. Results

Figure 2 shows transaxial, coronal, and sagittal slices of the denoised results using different methods. 30-sec image is used as the input noisy image and denoised images generated by 2D N2C, 2D N2N, 3D N2C, and 3D N2N are compared with the 600-sec image. Table 1 shows the PSNR and SSIM of the input and denoised ouputs with reference to the 600-sec image.

The overall denoised results gave remarkable performance for enhancing the quality of the PET image. We can see that the N2N method gave comparable PSNR and SSIM to N2C. Also the result for 3D convolution gave higher PSNR and SSIM than 2D convolution, resulting a denoised image with better details.

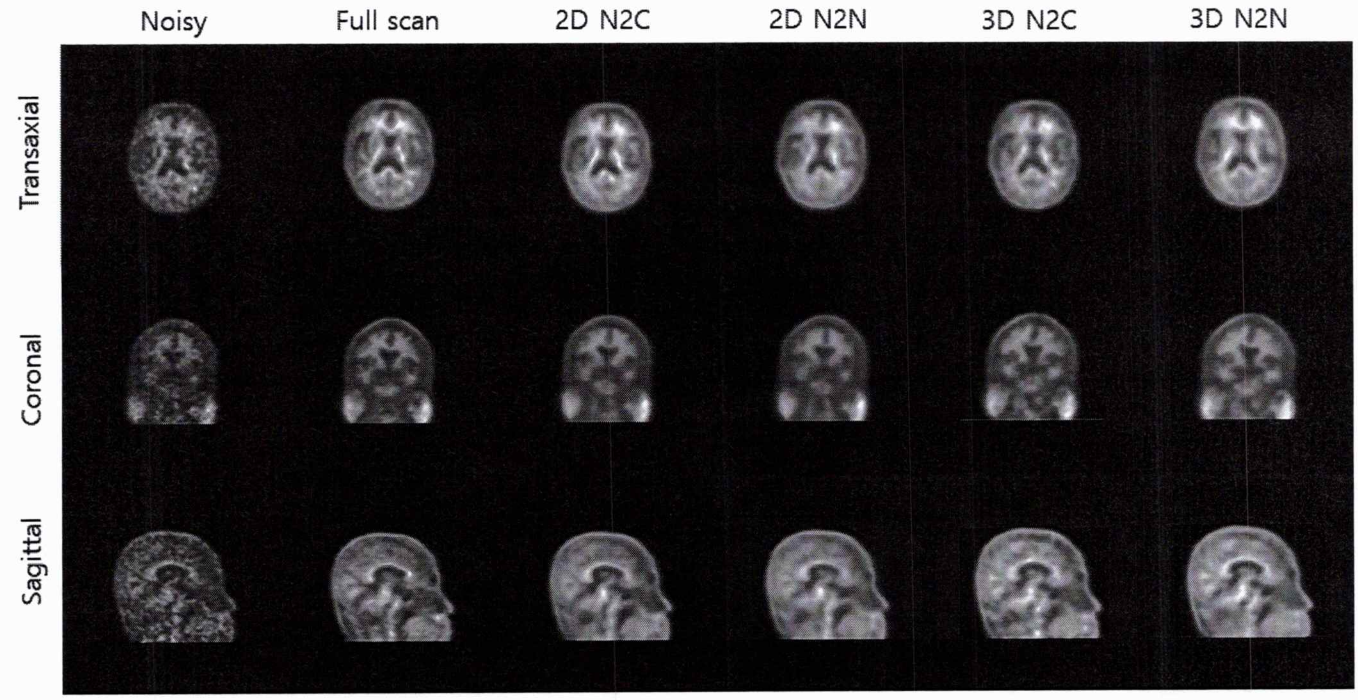

Figure 2. Transaxial, coronal, sagittal view of input noisy image of 30-sec scan time and denoised images using 2D N2C, 2D N2N, 3D N2C, and 3D N2N.

#	Input		2D Noise2Clean		2D Noise2Noise		3D Noise2Clean		3D Noise2Noise	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	29.11	0.9420	31.99	0.9685	32.09	0.9653	31.95	0.9854	31.68	0.9845
2	32.26	0.9555	34.83	0.9768	34.28	0.9681	34.89	0.9951	34.50	0.9795
3	30.12	0.9438	33.06	0.9701	32.76	0.9623	33.08	0.9934	33.18	0.9918
4	29.03	0.9300	32.68	0.9656	32.30	0.9573	32.59	0.9811	32.54	0.9926
5	29.77	0.9409	33.16	0.9704	32.85	0.9626	33.18	0.9888	33.34	0.9914
6	27.33	0.9187	30.48	0.9514	30.59	0.9478	31.39	0.9690	31.91	0.9913
7	30.62	0.9414	33.75	0.9716	33.27	0.9631	33.93	0.9952	34.04	0.9927
Mean	**29.75**	**0.9389**	**32.85**	**0.9678**	**32.59**	**0.9609**	**33.00**	**0.9869**	**33.03**	**0.9891**

Table 1. PSNR and SSIM of input and denoised output images for 7 test patients

5. Discussion

In this study, transaxial, coronal, sagittal slices were used together to probe 3D contexts when training 2D U-Net. However, training the slices from different views separately would give a better result in denoising the images. Furthermore, tri-planar representation where three views from a voxel are regarded as the three channels of 2D input can also be considered.

When using CNN for PET image denoising, intensity normalization and region of interest (ROI) should be considered together. In PET images, the original value for each voxel has a wide dynamic range reaching thousands. Intensity normalization for the pixel value of PET image is needed for effective training of the network. In this study, the 99^{th} percentile value was used to divide the pixel value of the images but other methods such as Z-score normalization can be considered together for better normalization.

Also, training the image only with the ROI would improve the performance of denoising. In this case, the brain would be the ROI and background should be excluded if possible. Furthermore, if the background of an image is included when measuring PSNR and SSIM, the quality of noisy image can be overestimated.

6. Conclusion

We compared 2D U-Net and 3D U-Net for Noise2Clean and Noise2Noise framework to reduce the noise in short scan time ^{18}F-Florbetaben brain PET images. The applied deep learning methods showed remarkable performance for reducing noise in PET images. Since Noise2Noise gave comparable results to Noise2Clean without using reference images, self-supervised methods can replace the traditional supervised method.

For future works, we will apply other convolution methods using transfer learning. To combine the advantages of both 2D and 3D approaches, we will use hybrid convolution such as H-Dense U-Net, which uses 2D pretrained networks with multi-slice inputs and 3D networks that are initialized randomly with volumetric inputs together for training. Additionally, networks using ACS convolution will be explored. ACS convolution is a 2D-to-3D transfer learning method which natively performs 3D representation learning while utilizing the pretrained weights on 2D datasets. Comparing hybrid and ACS convolution with 2D and 3D convolutions will help us how to improve the quality of 3D PET denoising. Moreover, we will apply advanced self-supervised methods such as Noiser2Noise to enhance the performance of denoising and generate sharper and detailed images.