Zero-dose PET Reconstruction with Missing Input by U-Net with Attention Modules

Jiahong Ouyang
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
jiahongo@stanford.edu

Kevin T. Chen
Department of Radiology
Stanford University
Stanford, CA 94305
ktchen@stanford.edu

Greg Zaharchuk
Department of Radiology
Stanford University
Stanford, CA 94305
gregz@stanford.edu

Abstract

Positron emission tomography (PET) is a widely used molecular imaging technique with many clinical applications. To obtain high quality images, the amount of injected radiotracer in current protocols leads to the risk of radiation exposure in scanned subjects. Recently, deep learning has been successfully used to enhance the quality of low-dose PET images. Extending this to "zero-dose," i.e., predicting PET images based solely on data from other imaging modalities such as multimodal MRI, is significantly more challenging but also much more impactful. In this work, we propose a attention-based framework that uses multi-contrast MRI to reconstruct PET images using the most commonly-used radiotracer, 18F-fluorodeoxyglucose (FDG), a marker of metabolism. We also introduce an input dropout training strategy to handle possible missing MRI contrasts. We evaluate our methods on a dataset of patients with brain tumors, showing the ability to create realistic and clinically-meaningful FDG brain PET images with low errors compared with full-dose ground truth PET images.

1 Introduction

Positron emission tomography (PET) is a widely used imaging technique in many clinical applications including tumor detection [1] and neurological disorder diagnosis [2]. To obtain high quality images, the amount of injected radiotracer in current protocols leads to the risk of radiation exposure in scanned subjects. Moreover, PET is expensive and not offered in the majority of medical centers in the world. On the contrary, MRI is a widely available and non-invasive technique. Therefore, it is of great value to achieve zero-dose PET reconstruction, meaning synthesizing high-resolution and clinically accurate PET images solely from other modalities, especially MRI.

With the boom of deep learning, there are some success on low-dose PET reconstruction: using 1% dose PET assisting with MRI to generate standard-dose PET [3]. However, whether it is possible to achieve zero-dose PET reconstruction is still an open question. There are limited existing works on it: Li et al. attempted to synthesize amyloid PET from MRI inputs using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [4], but the outputs were of low resolution and the synthesis had obvious deviation on AD subjects [5]. Wei et al. designed a patch-based coarse-to-fine architecture to generate myelin PET from multi-contrast MRI [6], but the model ignored some global information due to the patch-based design and treated all regions of the brain equally, which might cause errors in the disease-specific regions. Guo et al. [7] applied a U-Net to synthesize oxygen-15 water PET cerebral blood flow by structural and functional MRI input. The method stacked 8 MRI contrasts as input, and missing any of them can lead to inaccurate prediction.

In this work, on top of the U-Net, we design (1) a symmetry-aware spatial attention module to generate spatial-wise attention maps that enhance the abnormal region of the brain; (2) a special

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

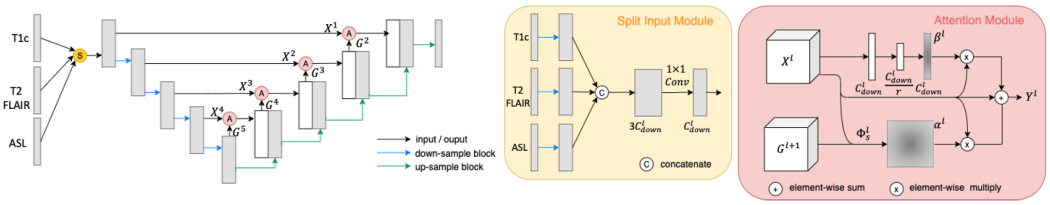

Figure 1: The overview of the proposed method.

split-input module with channel-wise attention to enhance the most important input contrasts; (3) a random input dropout training strategy to handle missing MRI contrasts. We evaluate the method on a private dataset including 38 Glioblastoma (GBM) cases. The results indicate the effectiveness of the proposed method.

2 Method

As shown in figure 1, a 2D U-Net works as the backbone, on which spatial-wise and channel-wise attention modules were added. We combine the attention modules on the short-cut connection of the U-Net. For the l-th layer, the weighted shortcut $\mathbf{Y}^l = (1 + \alpha^l + \beta^l)\mathbf{X}^l$, where α^l and β^l are spatial and channel-wise attention maps respectively, while $\mathbf{X}^l$ is the feature from the downstream part of the U-Net.

Symmetry-aware spatial-wise attention module (SSA). For the l-th layer, SSA takes features with a larger perception field from the lower upstream layer $\mathbf{G}^{l+1}$ (gate signal) as the "guide" for learning attention for the detailed features from the downstream layer $\mathbf{X}^l$ (input signal). Moreover, given that the normal brain is roughly symmetric, a radiologist will tend to compare the left and right sides of the brain and take advantage of asymmetry to help them find any abnormalities. We utilize the symmetry property by flipping the gate signal $\mathbf{G}^l$. More specifically, the SSA-map can be formulated as:

$$\alpha^l = Softmax(\mathbf{W}_s^l \cdot ReLU(\mathbf{W}_{sx}^l \cdot \mathbf{X}^l + Upsample(\mathbf{W}_{sg1}^l \cdot \mathbf{G}^l + \mathbf{W}_{sg2}^l \cdot \hat{\mathbf{G}}^l)) \tag{1}$$

where $\hat{\mathbf{G}}^l = abs(\mathbf{G} - flip(\mathbf{G}))$. $\mathbf{W}_{s*}^l$ are weights of linear layers.

Channel-wise attention module (CA). Learning from [8], we combine the squeeze-and-excitation mechanism in our model as the channel-wise attention module. It factors out the spatial dependency by global average pooling to learn a channel specific value that can be used to re-weight the feature maps and emphasize the useful channels. The CA-map can be note as:

$$\beta^l = Sigmoid(\mathbf{W}_{c2}^l \cdot ReLU(\mathbf{W}_{c1}^l \cdot AvgPool(\mathbf{X}^l))) \tag{2}$$

where $\mathbf{W}_{c*}^l$ are weights of linear layers.

Split input module with CA (SI). For multi-contrast inputs scenario, channels may include redundant information. Some channels are more important than others (e.g. ASL), and for many cases some of the input channels do not have strong signals (e.g. ASL). We design an input module that first splits each channel for a separate convolutional layer, and then combine them with CA module before feed into U-Net.

Random input dropout. When using multi-modality (contrast) inputs, it is common that some modalities are missing. Thus it is significant for the model to generate robust results based on incomplete information. We introduce a random input dropout training strategy [9] such that, for n modalities, $\sum_{i=0}^{n-1} \binom{n}{i} = 2^n - 1$ types of dropout scenario with selected probabilities are simulated during the training.

3 Results

Dataset and data preprocessing. Our dataset consists of 38 GBM cases with paired FDG PET and MRI (T1 with contrast (T1c), T2-FLAIR, and ASL) exams. For image pre-processing, MRI

	U-Net (U) [10]	Att. U-Net [11]	Proposed (U+SSA)	Proposed (U+SSA+CA)	Proposed (U+SSA+CA+SI)
PSNR	26.979±9.500	28.773±12.312	29.246±13.138	29.491±12.562	**29.677**±12.907
SSIM	0.799±0.007	0.850±0.006	0.861±0.006	0.860±0.005	**0.868**±0.005
RMSE	0.303±0.007	0.254±0.010	0.243±0.008	0.240±0.006	**0.226**±0.005

Table 1: Quantitative results.

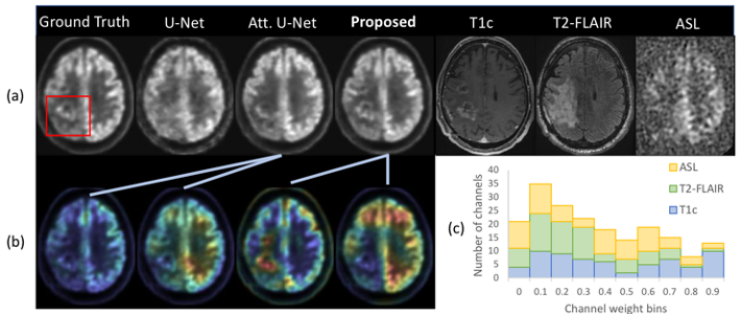

Figure 2: Qualitative results: (a) Inputs and results from each model. (b) Spatial attention maps from the middle two layers. (c) Channel attention histogram from the split-input module.

images were co-registered to PET images using SPM12. To better utilize the symmetry of the brain, PET and registered MRI images were further normalized to a standard template, with the size of $189 \times 157 \times 156$. Then, the volumes were normalized by the mean of the non-zero regions. Flipping along the X axis was used to avoid overfitting. Case-wise 5-fold validation was adopted.

Results without missing contrasts. Quantitatively, the reconstructed image quality was assessed by three metrics: peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and root mean square error (RMSE). The statistics for each slice (with the top and bottom 20 slices removed) were averaged to obtain the metrics. We also compared our results with U-Net [10] and attention U-Net (Att. U-Net) [11]. Meanwhile, we did an ablation study gradually adding each component. As shown in Table 1, the proposed method improved 10.00% in PSNR, 8.63% in SSIM, and 25.41% in RMSE comparing to the baseline U-Net, suggesting the effectiveness of SSA and CA as well as SI module.

Qualitatively, a typical example is shown in Figure 2-(a). All three methods performed similarly on the normal regions of the brain. However, the proposed method generated images with more accurate pathological features on the tumor region. We show the attention maps for the middle two layers in Figure 2-(b). The attention map from the proposed method precisely highlighted the tumor region, while maps from the attention U-Net only had weak signals, which might lead to the incorrect output. Figure 2-(c) presents the channel-wise attention weight distributions from the split input module. In this example, T1c provides the most information for the tumor and some parts of ASL have weak signal, which is consistent with the distribution that T1c is the most highly-weighted channel, while other two channel weights are lower.

Results with missing contrasts. Figure 3 demonstrates results from the proposed model with random input dropout strategy. Though not as good as model without input dropout, it achieved acceptable results except for ASL only as input, as ASL has weak signal and does not provide fine-grained structural information.

4 Conclusion

In this work, we proposed a U-Net based network with symmetry-aware spatial attention module to capture the abnormal region of the brain, and split-input with channel-wise attention module to enhance the most important contrasts, as well as a random dropout training strategy to handle missing input contrasts. The results on the GBM dataset illustrated the possibility of zero-dose PET reconstruction using multi-contrast MRI. Future work includes adopting advanced backbone

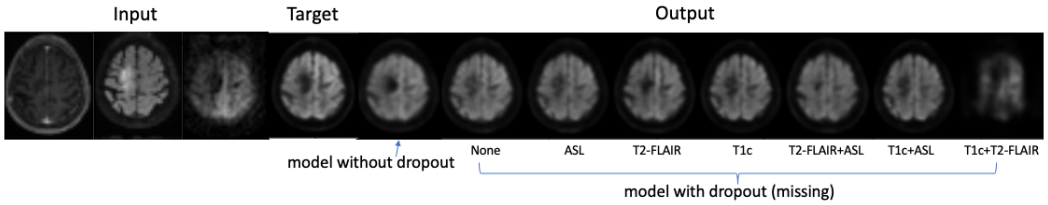

Figure 3: Results with missing contrast. For outputs from model with input dropout, missing: No, ASL, T2-FLAIR, T1c, T2-FLAIR+ASL, T1c+ASL, T1c+T2-FLAIR.

network, designing better way for missing contrasts, and evaluating the reconstructed image by tumor segmentation.

Broader Impact Statement

From the methodology aspect, the proposed modules can be easily plugged into any other network structure, even those not based on a U-Net-like segmentation/reconstruction network. SSA is a significant contribution that takes advantage of how radiologists use symmetry to guide their attention; this will be particularly applicable to neuroimaging. SI can be used whenever multiple input channels are available and is a more intelligent way of combining them than simple channel-stacking. Random dropout training strategy is a broadly applicable way to handle missing inputs, which is important for real-world applications.

From the applications aspect, many sites do not have access to advanced imaging such as PET. Developing a method than can synthesize the most common PET imaging examination (FDG) from multimodal MRI is an excellent starting point, given the close ties between metabolism (FDG PET) and blood flow (ASL MRI). Further work can explore the potential to use this or similar methods to other PET radiotracers, such as the dementia imaging agents amyloid and tau. Removing the need to dose the patient with radiation would allow this technology to be used in more vulnerable patients, such as pregnant women and children. Zero-dose PET will "level the playing field" between the few centers that can provide advanced technologies like PET and the many sites that cannot, democratizing imaging and acting as a force for more equitable medical care.

References

[1] Ken Ono, Reiji Ochiai, Tsuyoshi Yoshida, Mami Kitagawa, Junichi Omagari, Hisashi Kobayashi, and Yasuyuki Yamashita. The detection rates and tumor clinical/pathological stages of whole-body fdg-pet cancer screening. *Annals of nuclear medicine*, 21(1):65–72, 2007.

[2] Karl Herholz, Eric Salmon, D Perani, JC Baron, Vjera Holthoff, L Frölich, P Schönknecht, K Ito, Rüdiger Mielke, Elke Kalbe, et al. Discrimination between alzheimer dementia and controls by automated analysis of multicenter fdg pet. *Neuroimage*, 17(1):302–316, 2002.

[3] Kevin T Chen, Enhao Gong, Fabiola Bezerra de Carvalho Macruz, Junshen Xu, Athanasia Boumis, Mehdi Khalighi, Kathleen L Poston, Sharon J Sha, Michael D Greicius, Elizabeth Mormino, et al. Ultra–low-dose 18f-florbetaben amyloid pet imaging using deep learning with multi-contrast mri inputs. *Radiology*, 290(3):649–656, 2018.

[4] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005.

[5] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–312. Springer, 2014.

[6] Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Nicholas Ayache, Bruno Stankoff, and Olivier Colliot. Learning myelin content in multiple sclerosis from multimodal

mri through adversarial training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 514–522. Springer, 2018.

[7] Jia Guo, Enhao Gong, Audrey P Fan, Maged Goubran, Mohammad M Khalighi, and Greg Zaharchuk. Predicting 15o-water pet cerebral blood flow maps from multi-contrast mri using a deep convolutional neural network with evaluation of training cohort bias. *Journal of Cerebral Blood Flow & Metabolism*, page 0271678X19888123, 2019.

[8] Qiaoying Huang, Xiao Chen, and Mariappan Nadar. End-to-end segmentation with recurrent attention neural network. *arXiv preprint arXiv:1812.02068*, 2018.

[9] Darvin Yi, Endre Grøvik, Michael Iv, Elizabeth Tong, Kyrre Eeg Emblem, Line Brennhaug Nilsen, Cathrine Saxhaug, Anna Latysheva, Kari Dolven Jacobsen, Åslaug Helland, et al. Mri pulse sequence integration for deep-learning based brain metastasis segmentation. *arXiv preprint arXiv:1912.08775*, 2019.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[11] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53:197–207, 2019.