Paper Reading No.2

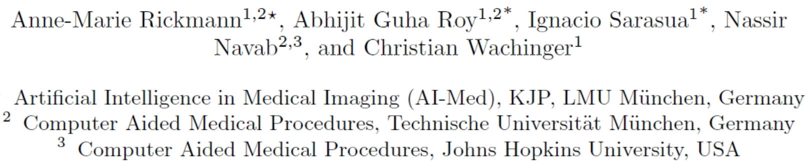
‘Project & Excite’ Modules for Segmentation of Volumetric Medical Scans

Sheng Lian

June 2019

1 Brief Paper Intro

- ***Paper ref:*** MICCAI 2019, https://arxiv.org/abs/1906.04649
- ***Authors:*** See Figure 1
- ***Paper summary:*** Inspired by the idea of SENet [1], this paper propose ‘Project & Excite’ (PE) modules that can deal well with 3D medical image segmentation. Instead of ’squeeze and excite’ steps as SENet do, this paper choose the ’project and excite’ way, which can retain more spatial information of medical volumetric images.
- ***Reading motivation:*** Papers of MICCAI-2019 are gradually released on arxiv. 3D medical image segmentation methods still have many drawbacks, whether it is 2D based method or 3D based method. This paper provides a new way to solve this problem.



Anne-Marie Rickmann[1,2⋆], Abhijit Guha Roy[1,2*], Ignacio Sarasua[1*], Nassir Navab[2,3], and Christian Wachinger[1]

[1] Artificial Intelligence in Medical Imaging (AI-Med), KJP, LMU München, Germany
[2] Computer Aided Medical Procedures, Technische Universität München, Germany
[3] Computer Aided Medical Procedures, Johns Hopkins University, USA

Figure 1: authors’ brief intro.

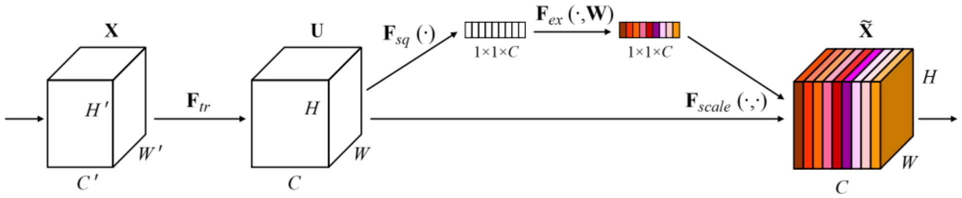


Figure 2: A Squeeze-and-Excitation block.

2 Backgrounds

For 3D medical image segmentation:

- 2D-CNN-based methods segment 3D medical scans slice-wise, and in this way the contextual information from adjacent slices remains unexplored.
- 3D-CNN-based methods have high computational complexity and are susceptible to the problem of over-fitting.

3 'Squeeze and Excite' Recap

SENet [1] is the winner of the last ImageNet competition on image classification mission. The core of this model is the squeeze and excitation operation, which explicitly model the interdependencies between CNN channels and adaptively recalibrate CNN feature map. In this way, SENet can selectivity enhance useful features and suppress less useful ones. SE block is flexible and can be easily applied to former CNN architectures, including VGGNet, Inception, ResNet, et al. Figure 2 briefly shows the main structure of SE-block. The procedure of 'squeeze and excite' operation can be summarized as follows:

- **Squeeze:** using global average pooling to generate channel-wise statistics. The c-th element of z(which can be regarded as weight vector) is calculated by Eq 1.

$$z_c = \mathbf{F}_{sq}\left(\mathbf{u}_c\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

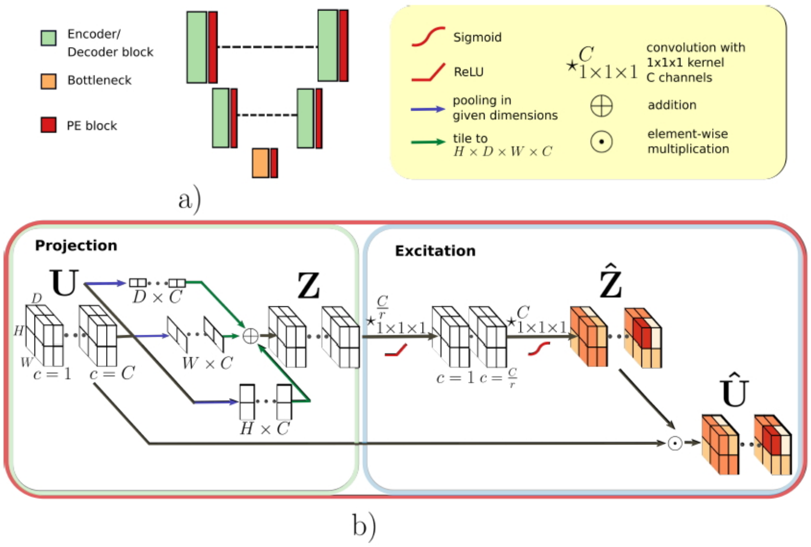


Figure 3: a): typical encoder/decoder based F-CNN architecture with PE blocks placed after each block. b): Illustration of the proposed 'Project & Excite' block. Projection operation, with the 3 different pooling operations and Excitation operation with 2 convolutional layers and recalibration of the feature map.

- **Excitation:**Learning $W \in \mathbb{R}^{c_2 \times c_2}$ to explicitly model channel association. This can be calculated by Eq 2

  $$\hat{\mathbf{z}} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma\left(\mathbf{W}_2 \delta\left(\mathbf{W}_1 \mathbf{z}\right)\right) \tag{2}$$

  with δ denoting ReLU, and σ the sigmoid layer. W_1 and W_2 are the weights of the fully-connected layers.

- **Scale:** Reweighting the feature maps. The output of the SE block is defined by a channel-wise multiplication of U with $\hat{\mathbf{z}}$.

  Finally in conclusion, The $c_t h$ channel of $\hat{U}$ is defined as:

  $$\hat{\mathbf{u}}_c = \mathbf{F}_{ex}\left(\mathbf{F}_{sq}\left(\mathbf{u}_c\right)\right) \mathbf{u}_c = \hat{z}_c \mathbf{u}_c \tag{3}$$

4 3D 'Project & Excite' Module

The 2D SE block can be easily transformed to 3D 'squeeze and excite' module through adding dimension. However, a volumetric input of large size holds relevant spatial information which might not be properly captured by a global pooling operation. Here, authors replace the spatial squeeze operation with projection operation. This follows the excite operation, which is same as SE block do.

The architecture details of PE block is illustrated in Figure 3. The projection operations are done by average pooling defined as:

$$\mathbf{z}_{h_c}(i) = \mathbf{F}_{pr_H}(\mathbf{u}_c) = \frac{1}{W}\frac{1}{D}\sum_{j=1}^{W}\sum_{k=1}^{D}\mathbf{u}_c(i,j,k), \quad i \in \{1,\ldots,H\} \tag{4}$$

$$\mathbf{z}_{w_c}(j) = \mathbf{F}_{pr_W}(\mathbf{u}_c) = \frac{1}{H}\frac{1}{D}\sum_{i=1}^{H}\sum_{k=1}^{D}\mathbf{u}_c(i,j,k), \quad j \in \{1,\ldots,W\} \tag{5}$$

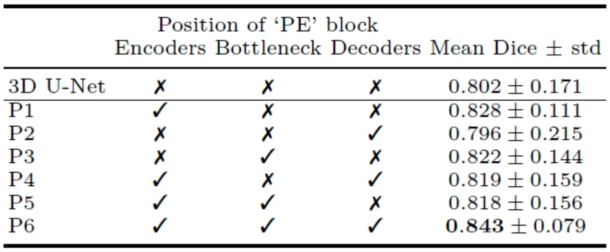
$$\mathbf{z}_{d_c}(k) = \mathbf{F}_{pr_D}(\mathbf{u}_c) = \frac{1}{H}\frac{1}{W}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbf{u}_c(i,j,k), \quad k \in \{1,\ldots,D\} \tag{6}$$

The outputs z_c are tiled to the shape H * W * D * C and added to obtain Z, which is then fed to the excitation operation $\mathbf{F}_{ex}(\cdot)$. The following operation is similar to SE block, where the FC layers are replaced by convolutional layers with kernel size 1*1*1.

5 Experiment

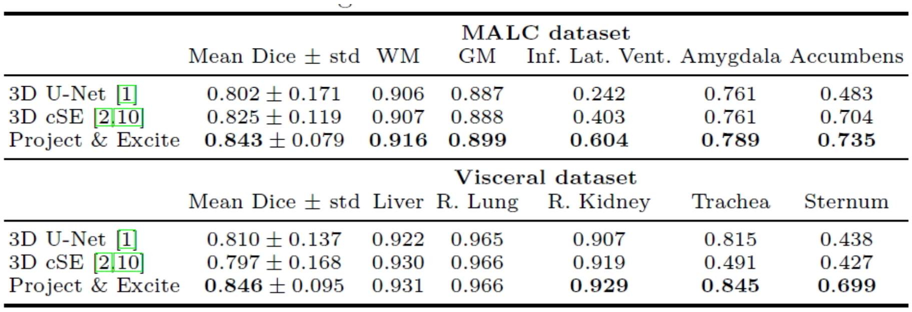
Authors choose the task of whole-brain segmentation (MALC dataset [3]) and whole-body segmentation (Visceral dataset [2]) for experiment. Both the two datasets only have limited scans($\tilde{2}0$).

Authors choose 3D U-Net as backbone model (demonstrated in Figure 3(a)). Here, authors discuss where the PE block should be placed in 3D UNet in Figure 4. The results show that the model perform best when the PE blocks are located in all the encoders, decoders and bottleneck, as is indicated in Figure 3(a).



	Position of 'PE' block			
	Encoders	Bottleneck	Decoders	Mean Dice ± std
3D U-Net	✗	✗	✗	0.802 ± 0.171
P1	✓	✗	✗	0.828 ± 0.111
P2	✗	✗	✓	0.796 ± 0.215
P3	✗	✓	✗	0.822 ± 0.144
P4	✓	✗	✓	0.819 ± 0.159
P5	✓	✓	✗	0.818 ± 0.156
P6	✓	✓	✓	**0.843** ± 0.079

Figure 4: Mean Dice score on MALC dataset due to placement of 'PE' blocks within 3D U-Net architecture..



	MALC dataset					
	Mean Dice ± std	WM	GM	Inf. Lat. Vent.	Amygdala	Accumbens
3D U-Net [1]	0.802 ± 0.171	0.906	0.887	0.242	0.761	0.483
3D cSE [2,10]	0.825 ± 0.119	0.907	0.888	0.403	0.761	0.704
Project & Excite	**0.843** ± 0.079	**0.916**	**0.899**	**0.604**	**0.789**	**0.735**
	Visceral dataset					
	Mean Dice ± std	Liver	R. Lung	R. Kidney	Trachea	Sternum
3D U-Net [1]	0.810 ± 0.137	0.922	0.965	0.907	0.815	0.438
3D cSE [2,10]	0.797 ± 0.168	0.930	0.966	0.919	0.491	0.427
Project & Excite	**0.846** ± 0.095	0.931	0.966	**0.929**	**0.845**	**0.699**

Figure 5: Comparison of 3D U-Net with 3D cSE and our proposed PE block. Mean Dice scores for selected classes of MALC and Visceral datasets. In the top table WM stands for white matter and GM for grey matter. In the bottom table L. stands for left and R. for right.

Also, Figure 5 shows the result comparison of PENet, the baseline 3D U-Net, and SENet. The results show the effectiveness of Project & Excite operation.

6 My thoughts

- As the champion of the last ImageNet classification mission, SENet has unparalleled advantages in feature extraction and feature selection. Adopting it to the task of 3D medical image segmentation can bring great advantages.
- Authors choose to replace squeeze with projection, this operation can combine spatial and channel context for recalibration. This step acts

as a bridge between 2D and 3D, and is rather interesting. The result comparison show the effectiveness of this operation.

References

[1] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[2] Oscar Jimenez-del Toro, Henning Müller, Markus Krenn, Katharina Gruenberg, Abdel Aziz Taha, Marianne Winterstein, Ivan Eggel, Antonio Foncubierta-Rodríguez, Orcun Goksel, András Jakab, et al. Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: Visceral anatomy benchmarks. *IEEE transactions on medical imaging*, 35(11):2459–2475, 2016.

[3] B Landman and S Warfield. Miccai 2012 workshop on multi-atlas labeling. In *Medical image computing and computer assisted intervention conference*, 2012.