**Swin-Unet Unet-like Pure Transformer for Medical Image Segmentation**

* Although CNN has achieved excellent performance, it cannot learn global and long-range semantic information interaction well due to the locality of convolution operation.
* Existing medical image segmentation methods mainly rely on fully convolutional neural network (FCNN) with U-shaped structure.
* Swin-Unet consists of encoder, bottleneck, decoder and skip connections. The basic unit of Swin-Unet is Swin Transformer block.
* Different from the conventional multi-head self-attention (MSA) module, Swin transformer block is constructed based on shifted windows.
* Since Transformer is too deep to be converged, only two successive Swin Transformer blocks are used to constructed the bottleneck to learn the deep feature representation. In the bottleneck, the feature dimension and resolution are kept unchanged.
* The Swin-Unet is achieved based on Python 3.6 and Pytorch 1.7.0. For all training cases, data augmentations such as flips and rotations are used to increase data diversity.
* Increase of model scale hardly improves the performance of the model, but increases the computational cost of the whole network.
* In order to leverage the power of Transformer, we take Swin Transformer block as the basic unit for feature representation and long-range semantic information interactive learning.
* Extensive experiments on multi-organ and cardiac segmentation tasks demonstrate that the proposed Swin-Unet has excellent performance and generalization ability.