Topic: Image segmentation by using U-net

Image segmentation is a classic problem in computer vision research, has become a focus in the field of image understanding, image segmentation is the first step of image analysis, is the basis of computer vision, is an important part of image understanding, but also one of the most difficult problems in image processing since its complexity. It is affected by many aspects, including noise, low contrast, illumination, and irregularity of object boundaries. Image segmentation is usually used to locate objects and boundaries (lines, curves, etc.) in an image. More precisely, image segmentation is the process of assigning a label to each pixel in the image so that pixels with the same label have certain characteristics.

Image segmentation is related to high-dimensional data. The image is composed of different colours. We can understand the pixel as the smallest unit of the image. The pixel is a value used to control the colour of such a smallest unit. The simplest image is a grayscale image, which is actually a black and white image. However, black is also divided into grades. For example, white can be understood as the lightest black, so white is defined as 0, black is defined as 255, and the rest of different degrees of grey are between 0 and 255. For example, a 10*10 grayscale image is an image composed of 10 pixels long and 10 pixels wide, with a total of 100 pixels, and each pixel value is between 0 and 255.

As for the image segmentation technology, because of the importance and difficulty of the problem itself, the image segmentation problem has attracted a lot of researchers to make great efforts since the 1970s. Although so far, there is no universal perfect image segmentation method, but for the general law of image segmentation is basically reached a consensus, has produced a considerable number of research results and methods. This report gives an introduction to U-Net and several improved versions.

In image segmentation tasks, especially medical image segmentation, U-Net is undoubtedly one of the most successful methods, which is our first chosen research paper: U-Net: Convolutional Networks for Biomedical Image Segmentation [1]. In this article, the authors proposed a network and training strategy that relies on the powerful use of data augmentation to make more effective use of available annotated samples. This method was proposed at the MICCAI conference in 2015 and has now reached more than 25,000 references. The encoder (down-sampling)-decoder (up-sampling) structure and skip connection adopted are a very classic design method.

The structure of U-Net is shown in the figure below. The left side can be regarded as an encoder, and the right side can be regarded as a decoder. The encoder has four sub-modules. Each sub-module contains two convolutional layers. After each sub-module, there is a down-sampling layer implemented by max pool. The resolution of the input image is 572x572, and the resolutions of modules 1-5 are 572x572, 284x284, 140x140, 68x68 and 32x32 respectively. Since the convolution uses the valid mode, the resolution of the next sub-module here

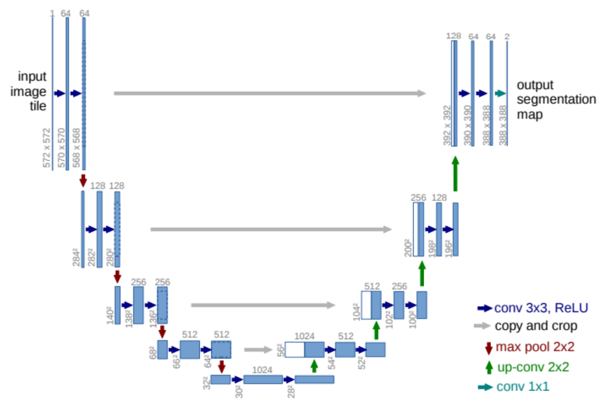

Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

is equal to (resolution of the previous sub-module -4)/2. The decoder contains four sub-modules, and the resolution is sequentially increased by up sampling until it is consistent with the resolution of the input image (because the convolution uses the valid mode, the actual output is smaller than the input image). The network also uses a skip connection to connect the up-sampling result with the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder.

3D U-Net is a simple extension of U-Net, which is introduced from our second chosen research paper: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [2]. It applied to 3D image segmentation; the structure is shown in the figure below. Compared with U-Net, this network only uses three down sampling operations, and batch normalization is used after each convolutional layer, but both 3D U-Net and U-Net do not use dropout.

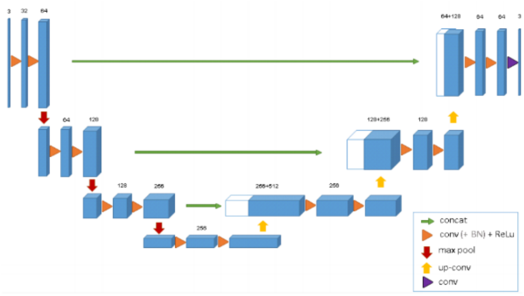

Fig. 2: The 3D u-net architecture. Blue boxes represent feature maps. The number of channels is denoted above each feature map.

The network structure, pre-processing, training and inference of U-Net may be different for different tasks. These choices are not independent of each other and affect the final outcome. The authors proposed NNU-Net (No New-Net) [3], which is an adaptive framework based on 2D and 3D U-Net. The author believes that we should not worry too much about the details of network design but should focus on other aspects that can improve the performance and generalization of the model. Then the authors carried out experiments and found that the results were surprisingly good. NNU-Net uses three relatively simple U-nets, and these U-nets are only minor modifications of the original U-Net, without various extensions (residual connection, Dense connection, and various attention mechanisms). NNU-Net can automatically adapt its architecture to a given image geometry, and more importantly, it thoroughly defines all the other steps around them.

Reference

[1] Ronneberger, O., Fischer, P. and Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. [online] arXiv.org. Available at: https://arxiv.org/abs/1505.04597.

[2] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv:1606.06650 [cs]*. [online] Available at: https://arxiv.org/abs/1606.06650 [Accessed 4 May 2021].

[3] Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. and Maier-Hein, K.H. (2019). No New-Net. *arXiv:1809.10483 [cs]*. [online] Available at: https://arxiv.org/abs/1809.10483? [Accessed 4 May 2021].