1. Processing

The image intensity values for all CT scans are truncated to range of [400, 2400] Hounsfield unit (HU) to remove the irrelevant details.

2. Network Architecture and train loss

The 2D network architecture for volumetric segmentation employed in this work is shown in Figure 1. We have an encoder path with five resolution steps on the left and a decoder path with three resolution steps on the right. The left part employs 2D convolution layers and residual blocks to learn the higher and higher representation features of medical images. We take convolution operation with stride of 2 to reduce the spatial resolution of feature map by half in encoder path. Inspired by FCN-8s, the right part decompresses the extracted higher features into finer and finer resolution through deconvolution layers. Apart from encoder path and decoder path, we imposed convolution block to bridge the skip connections between low-level features and high-level features. Moreover, there are three auxiliary loss layers and one main loss layers in our network. For the auxiliary loss layers, we apply deconvolution layers to upsample feature maps to be the same spatial resolution as input.

Binary cross entropy function is employed as loss function in this work, which is described as:

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log \hat{y}_i + (1-y_i)(1-\log \hat{y}_i)\right) \quad (2)$$

where y represents the ground truth, $\hat{y}$ denotes the predicted segmentation results, y_i and $\hat{y}_i$ indicates the label and predicted probability for voxel i, respectively. The overall loss is formulated as below:

$$L_{total} = L(y,\hat{y}_{main}) + \beta_1 L(y,\hat{y}_{aux1}) + \beta_2 L(y,\hat{y}_{aux2}) + \beta_3 L(y,\hat{y}_{aux3}) \quad (3)$$

Where $L(y,\hat{y}_{main})$, $L(y,\hat{y}_{aux1})$, $L(y,\hat{y}_{aux2})$, $L(y,\hat{y}_{aux3})$ denotes the main loss and three auxiliary loss, respectively. β_1, β_2, β_3 are the balanced weights and set as 0.2, 0.4, 0.8, respectively.

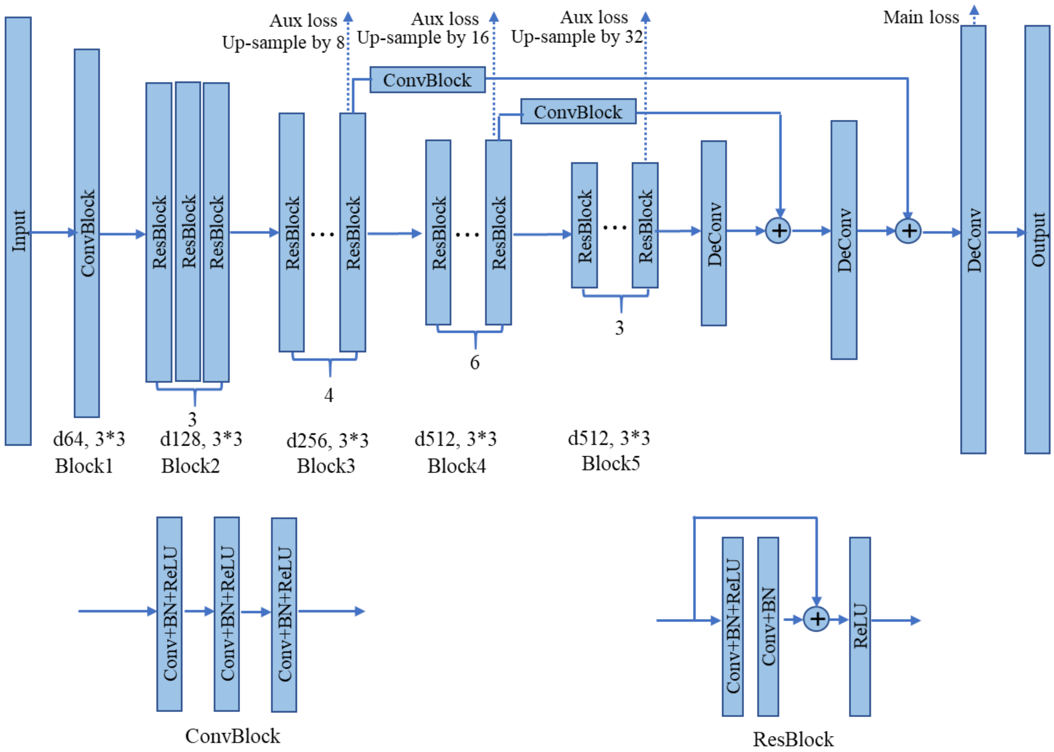

Figure 1. Illustration of 2D convolutional network architecture for volumetric segmentation used in this aneurysm segmentation.

2. Sample Balancing Strategy

We assume that there are m cases in training dataset, X_i denotes the i-th case, $i = 1, \cdots, m$, and there are n slices in X_i, $x_{i,j}$ denotes the j-th slices in X_i, $j = 1, \cdots, n$. As shown in Figure 2A, there are two type of slices in X_i: negative slices and positive slices. Generally, the positive slices are at the middle, denote as P_i, while the negative slices at both ends, denoted as $N_{i,1}$ and $N_{i,2}$, respectively. We assume that there are k positive slices in X_i, and $x_{i,l}$ is the first positive slice, so $N_{i,1} = \{\mathrm{x}_{i,1}, \cdots, x_{i,l-1}\}$, $P_i = \{x_{i,l}, \cdots, x_{i,l+k-1}\}$, $N_{i,2} = \{x_{i,l+k}, \cdots, x_n\}$. In medical volumetric images, adjacent slices have strong correlation, which is useful in information complementation, especially spatial information. In order to effectively extract the feature information of positive slices, in this work all positive slices and their adjacent continuous negative slices are selected in every CT scan (Figure 2A). We set the selected ration on negative slices in CT scan as r, $r \in [0,1]$. The selected negative slices are denoted as $SN_{i,1}$ and $SN_{i,2}$, and the number of slice in $SN_{i,1}$ and $SN_{i,2}$ are $n_{i,1}$ and $n_{i,2}$, respectively.

$$n_{i,1} = \lceil (l-1) \times r \rceil \quad (4)$$

$$n_{i,2} = \lfloor (n-l-k+1) \times r \rfloor \quad (5)$$

So $SN_{i,1} = \{x_{i,l-n_{i,1}}, \cdots, x_{i,l-1}\}$ and $SN_{i,2} = \{x_{i,l+k}, \cdots, x_{i,l+k+n_{i,2}-1}\}$. The selected samples in X_i are denoted as $SX_i^r = \{SN_{i,1}, P_i, SN_{i,2}\}$, therefore, the selected training dataset as $D^r = \{SX_i^r, \cdots, SX_n^r\}$. As mentioned above, we could get that if $0 \leq r_1 < r_2 \leq 1$, D^{r_2} is a subset of D^{r_1}. As shown in Figure 2B, we gradually reduce the negative slices in training dataset by changing r value, this indicates that the later selected training dataset is a subset of the previous one. As for the selection is between positive slices and negative slices, so the proposed method is named as positive-negative subset selection. For convenience, we used $L(\alpha, \beta)$ to regulate r, which represents the value of r reduced α after β epochs until $r = 0$.

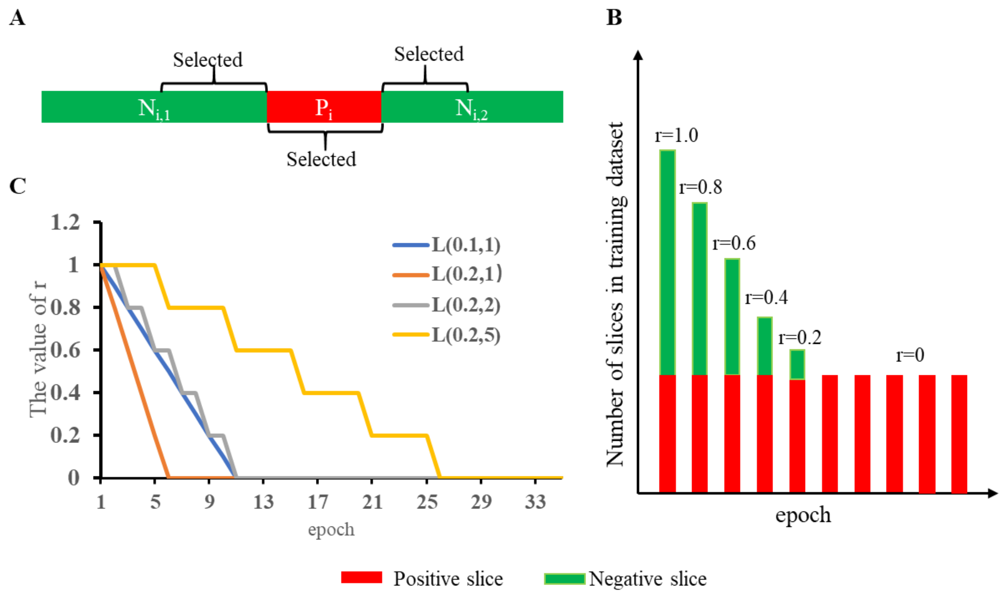

Figure 2. Example of subset sample selection method. A displays sample selection method on individual 3D CT-scan, B displays sample selection method in training process and C displays that how $L(\alpha, \beta)$ regulates r value. Red and green denote the positive slices and negative slice, respectively. And r represents the selected ration on negative slices.

3. Coarse-to-fine strategy

In aneurysm segmentation task, the coarse-to-fine strategy. And to improve the spatial feature extraction ability of network, 9 adjacent images are selected. We first get the coarse segmentation results in original image size (256*256). And the we crop the ROI (96*96) for the fine segmentation results. We use connected domain processing to remove noise (volume <20 pixels). At the same time, in order to improve the segmentation effect, we use classification network to determine whether the connected area is aneurysm, if the connected area is greater than 2.