Solutions for Airway segmentation

Yang Nan, Shiyi Wang, Guang Yang
National Heart and Lung Institute, Imperial College London, London, UK

Given an input raw data, i.e., a 3D volume $\mathcal{V} \in R^{W\times H\times Z}$, a patch sampling strategy is first presented to extract appropriate patches $\mathcal{V}_P \in R^{w\times h\times c}$ for training and validation. Then, a fuzzy attention neural network is proposed to segment the airway trees with the supervision of a comprehensive loss function (including a Dice loss, an airway continuity loss, and an accumulation mapping loss) on $\mathcal{V}_P$. During the training, the AF-score is developed to assess the model performance and save the best values of the trainable network parameters.

The proposed FANN is built based on the 3D U-Net, adding a novel fuzzy attention layer, deep supervision, and airway continuity and accumulation mapping (ACAM) loss. The size of the volume patches $\mathcal{V}_P$ is set as the same ratio as the median shape of the ground truth annotations in the training data. FANN includes 3D convolution layers, instance normalization (IN), LeakyReLu (LReLU), 3D transpose convolution layers, fuzzy attention layers, and sigmoid activation layers. It is of note that the proposed network has multiple outputs, including the main output and three low-level outputs. The final prediction is given from the main output, while the low-level outputs are collected for deep supervision through the auxiliary losses (losses calculated as the same way as that of the main output with different weights). When calculating these auxiliary losses, the prediction is upsampled to align with the size of ground truth masks.

Patch sampling:

1) The average size $S\ (z \times y \times x)$ of the 3D minimum bounding volume of ground truth annotations is first calculated. The patch size is set as the similar scale as S.
2) The centerline of manual annotation is extracted by skeletonization [1].
3) Overlapped sliding windows are adopted to extract image patches, mask patches, and centerline patches.

Fuzzy Attention Layer:

In addition to the "gradient vanishing" problem, one major concern with the sigmoid activation function is its sharpness, by which only a small interval can obtain the output value ranges between 0 and 1. Therefore, it is difficult to find a robust "boundary" in sigmoid activation that distinguishes whether the feature is relevant or not. Another issue is the monotonicity: similar to the raw intensity distribution of airway structures that has both negative and positive variations, the feature representation of airway regions should also have two side variations. However, to align with the sigmoid activation function, the 1×1 convolution layers must learn to shift the values of all the features of interest to a single side to obtain a positive response. In particular, there must exist a certain "threshold" in the feature representation reconstitution (usually accomplished by the 1×1 convolutional layers in the attention layer) to determine whether the region is important or not. Moreover, non-channel specifics of the current attention map assign the same "attention" coefficient to all the feature points along the channel-wise. Specifically, given a feature representation $\mathcal{F} \in R^{C\times W\times H\times D}$, the existing attention map is formulated as $\alpha \in R^{W\times H\times D}$, while all the feature representations along the channel wise C shares the same 'importance'. This mechanism is unreliable, since the feature representations in different channels are extracted by different convolution kernels; therefore, we advocate the attention map to be channel-specific.

Different from most previous studies that applied the fuzzy logic 'AND' to fuzzy sets to obtain fuzzy feature representations [2-4], our goal is to use the fuzzy membership function to learn the 'importance' of target feature representations. Therefore, we point that the information can be better preserved by applying fuzzy logic 'OR' to fuzzy sets while suppressing irrelevant features. Having two fuzzy sets $\widetilde{A}$ and $\widetilde{B}$, the fuzzy logic 'OR' is described as

$$f_{\widetilde{A}\cup\widetilde{B}}(y) = f_{\widetilde{A}} \vee f_{\widetilde{B}} \quad \forall y \in U, \tag{1}$$

where U is the universe of information and y is the element of the universe. To make the fuzzy logic 'OR' derivative, we modify it as

$$f_{\widetilde{A}\cup\widetilde{B}}(y) = \max(f_{\widetilde{A}}, f_{\widetilde{B}})\,. \tag{2}$$

Therefore, the fuzzy degree $f_j(X,\mu,\sigma) \in \Theta^{H\times W\times D}, \Theta \in [0,1]$ of the j-th channel can be obtained.

Loss function:

Overall, the L_{JCAM} can be summarised as

$$L_{JCAM} = \alpha L_{\mathcal{J}}(X,Y) + \beta L_C(X,Y_{CL}) + \varphi L_{CE}(X,Y)$$
$$\gamma L_{LAM}(X,Y) + \delta L_{nLAM}(X,Y), \quad (3)$$

where $\alpha, \beta, \gamma, \varphi, \delta$ are the weights of Jaccard loss $L_{\mathcal{J}}$, continuity loss L_C, cross-entropy loss L_{CE}, linear and nonlinear accumulation mapping loss L_{LAM}, L_{nLAM}, respectively.

With the ground truth centerlines Y_{CL}, the continuity can be evaluated by calculating the ratio of correct-predicted centerlines and ground truth centerlines, that is

$$\begin{aligned} L_C &= 1 - C \\ &= 1 - \frac{\sum(X \cdot Y_{CL})}{\sum Y_{CL}} \end{aligned} \quad (4)$$

without considering the airway branch size.

Denote $A \oplus \tau$ as the summation operation that sums the array $A \in R^{N\times C\times W\times H\times D}$ along τth channel. The LAM and nLAM are calculated by

$$L_{LAM} = \sum\nolimits_{\tau\in[W,H,D]} \ell_1(X\oplus\tau, Y\oplus\tau), \quad (5)$$

$$L_{nLAM} = \sum\nolimits_{\tau\in[W,H,D]} L_{\mathcal{J}}\big(tanh(X\oplus\tau), tanh(Y\oplus\tau)\big), \quad (6)$$

where $L_{\mathcal{J}}$ is the Jaccard loss defined as

$$\begin{aligned} L_{\mathcal{J}} &= 1 - \mathcal{J} \\ &= 1 - \frac{XY + \varepsilon}{X + Y - XY + \varepsilon}, \end{aligned} \quad (7)$$

Evaluation Metric:

Here we propose a continuity and completeness F-score (CCF-score) for the aforementioned purpose

$$CCF_s = (1 + \omega^2) \times \frac{J \times C}{\omega^2 \times J + C}, \quad (8)$$

where $\omega \in R$ [0,1] is the preference parameter, J and C are the Jaccard index and Continuity index as defined in Eq.(10) and Eq. (13), respectively. ω can be set larger (smaller) than 1 when C (J) is more important. With this novel F-measure based metric, the module that focuses on both continuity and completeness can be saved. Here we set $\omega = 0.5$.

Training Details:

Parameters of the proposed 3D-UNet are initialized with He-normal initialization. Randomized rotation (rotation degree ranges from -10° to 10°), randomized flip (up, down, left, right) were implemented to augment the dataset during training. All modules were trained on an NVIDIA RTX 3090 GPU for 200 epochs, with an initial learning rate of 1e^(-3) and a decay of 0.5 at the 20th, 50th, 80th. 110th and 150th epoch. For equal contribution of L_J, L_C, L_CE, the value of the hyper-parameters α,β,γ in L_JCAM was set to 1, and that of φ, δ was set to 0.3 to constraint the accumulation mapping loss (since it is calculated through three dimension). The ω in CCF-score was set to 0.9 to prevent excessive leakages.

Acknowledgements

It is of note that this method has been submitted to a journal.

[1] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-D medial surface axis thinning algorithms," *CVGIP: Graphical Models and Image Processing,* vol. 56, no. 6, pp. 462-478, 1994.

[2] C. Guan, S. Wang, and A. W.-C. Liew, "Lip image segmentation based on a fuzzy convolutional neural network," *IEEE Transactions on Fuzzy Systems,* vol. 28, no. 7, pp. 1242-1251, 2019.

[3] Y. Deng, Z. Ren, Y. Kong, F. Bao, and Q. Dai, "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Transactions on Fuzzy Systems,* vol. 25, no. 4, pp. 1006-1012, 2016.

[4] S. R. Price, S. R. Price, and D. T. Anderson, "Introducing fuzzy layers for deep learning," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019: IEEE, pp. 1-6.