
Deep Learning For 3D Kidney Vessel Segmentation

Mutian Hong
ShanghaiTech University
2022533109
hongmt2022@shanghaitech.edu.cn

Zhengxuan Wei
ShanghaiTech University
2022533077
weizhx2022@shanghaitech.edu.cn

Junhui Ge
ShanghaiTech University
2022533047
gejh2022@shanghaitech.edu.cn

Abstract

Kidney vessel segmentation in medical imaging is crucial for diagnosing renal diseases. Manual annotation is time-consuming and lacks scalability, while existing methods struggle with using spatial information effectively. We propose a 2.5D deep learning method which exploits the spatial information of adjacent slices and combines spatial information from multiple angles in the inference stage. We experimented with a variety of deep learning models, demonstrates superior performance in delineating complex vascular structures within high-resolution 3D scans.

1 Introduction

Automating the segmentation of kidney vessels is crucial for medical imaging, diagnosis, and treatment of renal diseases. Manual annotation is time-consuming and lacks scalability. Machine learning approaches can accelerate this process by automating segmentation, contributing to the completion and refinement of the VCCF and Human Reference Atlas (HRA).

Deep learning has shown success in medical image analysis, but kidney vessel segmentation presents challenges due to intricate structures and 3D scans. Traditional architectures like UNet[1] struggle with spatial information in 3D images. Existing 3D convolutional networks are computationally expensive.

To address these challenges, we propose a 2.5D convolutional neural network architecture leveraging adjacent layers as inputs in different channels. This approach efficiently incorporates spatial context to improve segmentation accuracy. During inference, we integrate information from multiple perspectives to enhance robustness, including results from three axes.

We rigorously evaluated our model in the Kaggle competition "Hacking the Human Vasculature in 3D"[2], comparing it with nnUnet. Our results show significant advancements in segmentation accuracy and performance, the visualization result is shown in figure 1.

2 Methods

2.1 Preprocessing

We partition the original input into fixed-size windows of height H and width W . Additionally, we include C layers above and below the window as extra channels, effectively capturing contextual information surrounding the target region. This process is represented as follows:

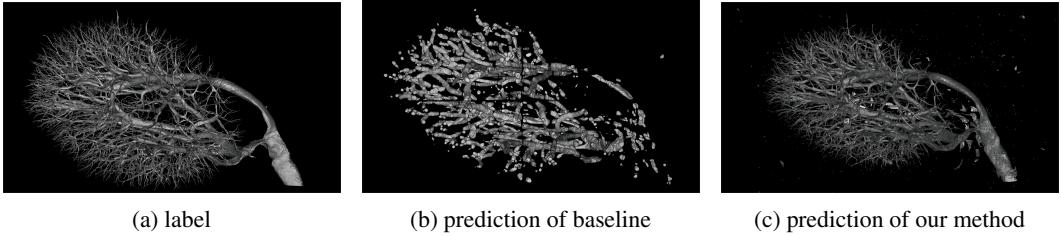


Figure 1: visualization result of kidney vessel

Let I_{raw} represent the raw input image with dimensions $H_{raw} \times W_{raw} \times D_{raw}$, where D_{raw} denotes the number of original channels. We partition I_{raw} into fixed-size windows of size $H \times W \times (2C + 1)$, where $2C + 1$ represents the total number of channels in each window.

The preprocessing pipeline includes the following transformations, denoted symbolically:

Window Partitioning: $I_{part} = \{I_1, I_2, \dots, I_n\}$, where each I_i represents a window of size $H \times W \times (2C + 1)$.

Random Transformations: Random transformations are applied to each window to augment the dataset and improve model generalization. Let $T(I_i)$ represent a random transformation applied to window I_i . Note that $T(I_i) = I_i$ for inference.

2.2 Swin Transformer

Swin Transformer[3] serve as the backbone architecture for our segmentation model, chosen for their ability to effectively capture long-range dependencies and contextual information in visual data, whose structure is shown in figure 2.

Let F_{Swin} denote the Swin Transformer function. It takes the preprocessed samples $T(I_i)$ obtained from the previous section as input and outputs processed features Y_i for downstream tasks:

$$Y_i = F_{Swin}(T(I_i))$$

The Swin Transformer’s hierarchical architecture efficiently models information at multiple scales with linear computational complexity relative to image size, facilitating the processing of high-resolution, large-scale image data. Shifted Windows enhance computational efficiency by confining self-attention computation to non-overlapping local windows while allowing cross-window connections, significantly improving scalability for processing large-scale visual data, such as high-resolution images for kidney vessel segmentation. Additionally, the Swin Transformer excels at fine-grained tasks, recognizing intricate details within images, making it theoretically advantageous for segmenting tiny blood vessels.

2.3 Training Method

2.3.1 Three Dimensional Rotation

Due to the scarcity of our data, we have obtained more training data by rotating the data at multiple angles within a three-dimensional space. We rotate the three-dimensional matrices in 3D space, using a constant interpolation method, and then extract the largest inscribed rectangular prism as our training data.

2.3.2 Loss Function

In binary segmentation tasks like ours, Binary Cross-Entropy (BCE) loss[4] is a common loss function. However, in our kidney vessel segmentation task, the number of positive samples is far fewer than the number of negative samples. BCE loss performs poorly when faced with such class imbalance issues. Dice loss[5] focuses more on positive samples. Therefore, it yields better results in this task.

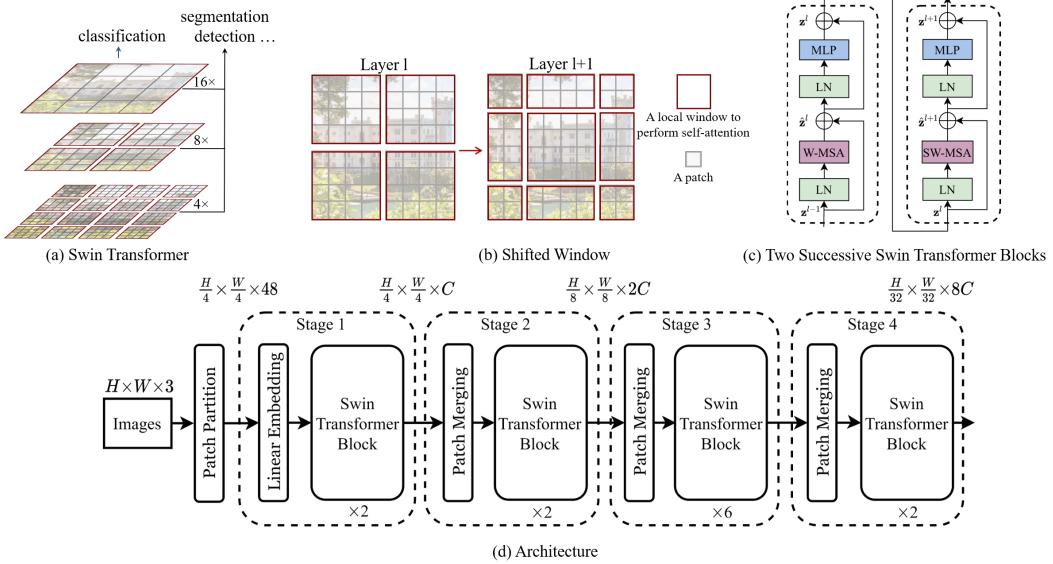


Figure 2: Swin Transformers Architecture

2.4 Inference Method

2.4.1 Three-Axis Inference

In the Three-Axis Inference approach, we extract I_x , I_y , and I_z , corresponding to the axial, sagittal, and coronal axes, respectively. Each set of samples undergoes independent prediction by the Swin Transformer model and tile splicing method, resulting in intermediate prediction maps: Y_x , Y_y , and Y_z . After obtaining predictions from all three axes, we integrate them to generate the final segmentation map, Y , using the mean operation:

$$Y = \text{mean}(Y_x, Y_y, Y_z)$$

This integration process combines information from multiple perspectives, enhancing the model's robustness and accuracy in capturing spatial structures and features across different orientations.

2.4.2 Drop Edge Pixels

The Tile Splicing Method introduce a potential complication that is commonly known as the "edge effect." The edge effect refers to the distortion that occurs at the boundary of each tile. Pixels located at the edges of these tiles suffer from a lack of contextual information. This absence of adjoining data points can lead to instability in the prediction outcomes for those particular edge pixels.

To address this issue, we employ a method to drop certain pixels along the edges during the stitching process. This effectively addressed the edge effect.

3 Experiments

3.1 Datasets

The dataset utilized in our experiment originates from the Kaggle competition titled "Hacking the Human Vasculature in 3D". Specifically, the competition dataset consists of high-resolution 3D images depicting multiple kidneys along with corresponding 3D segmentation masks delineating their vasculature. The acquisition of kidney images was facilitated through Hierarchical Phase-Contrast Tomography (HiP-CT)[6] imaging techniques. Notably, the dataset encompasses three distinct kidneys. As the testing set for the competition was not publicly available, we elected to utilize the second kidney from the training set as our experimental testing dataset. Moreover, training procedures were conducted exclusively on the remaining two kidneys.

Model	Loss	Three-Axis Infer	Edge Drop	Dice Score (%)	Surface Dice Score (%)
UNet	BCE	x	x	50.17	11.55
UNet	BCE	✓	x	61.66	12.35
UNet	Dice	✓	x	80.34	18.60
Swin	BCE	✓	x	78.99	60.45
Swin	Dice	✓	x	81.78	62.69
Swin	Dice	✓	✓	82.03	64.21
nnUNet	-	-	-	80.63	62.24

Table 1: Ablation experiments and comparison with other models

3.2 Evaluation Metrics

To assess the outcomes predicted by our model, we have adopted the Dice score (DSC)[7] and Surface Dice score (SDSC)[8] as our evaluative metrics.

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

X denotes the set of points of the predicted segmentation, and Y represents points of the ground truth set.

$$SDSC = \frac{2 \times |X_s \cap Y_s|}{|X_s| + |Y_s|}$$

We set the tolerance τ to 0. X_s means the set of surface points of the predicted segmentation, and Y_s means the set of surface points of the ground truth.

3.3 Experiment Result

The ablation experiments and comparison with other models is shown in Table 1.

analysis:(i) The model we used significantly outperforms the baseline UNet model and excels beyond nnUNet, the current cutting-edge model in medical image segmentation. (ii) For this task, due to its focus on positive samples, Diceloss outperforms BCEloss. (iii) Three-axis inference better utilizes spatial information, resulting in significant performance improvements for the model. Drop Edge Pixels addresses edge effects during inference and further optimizes the model.

3.4 Implementation Details

In the final deployment, we set $H = W = 224$, D (total number of channels) = $2C + 1 = 3$. We employed the following data augmentation techniques: random rotations, cropping, and the addition of Gaussian noise. During training, we set the batch size to 64 and the learning rate to 1×10^{-4} . The model was trained for 25 epochs on 4 * P40 GPUs, taking approximately 3 hours.

4 Conclusion

Our study presents a novel approach to kidney vessel segmentation in medical imaging using a 2.5D deep learning method. By leveraging spatial information from adjacent slices and integrating multiple perspectives during inference, our model demonstrates superior performance in delineating complex vascular structures within high-resolution 3D scans. Our experiments on the Kaggle competition dataset showcase the effectiveness of our approach, achieving a Dice score of 82.03% and a Surface Dice score of 64.21%. These results underscore the potential of our method for improving the automation and accuracy of kidney vessel segmentation tasks in medical imaging.

In practical deployment, our model offers scalability and efficiency, with implementation details optimized for real-world application. With further refinement and validation, our approach holds promise for enhancing diagnostic and treatment capabilities in renal diseases, paving the way for advancements in medical image analysis and healthcare technology.

References

- [1] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597, 2015.
- [2] Yashvardhan Jain, Katy Borner, Claire Walsh, Nancy Ruschman, Peter D. Lee, Griffin M. Weber, Ryan Holbrook, Addison Howard. (2023). SenNet + HOA - Hacking the Human Vasculature in 3D. Kaggle. <https://kaggle.com/competitions/blood-vessel-segmentation>
- [3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv preprint arXiv:2103.14030, 2021.
- [4] Ma Yi-de, Liu Qing, and Qian Zhi-Bai. Automated image segmentation using improved pcnn model based on cross-entropy. In Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004., pages 743–746. IEEE, 2004.
- [5] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, M. Jorge Cardoso. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In Lecture Notes in Computer Science, pages 240–248, Springer International Publishing, 2017. DOI: 10.1007/978-3-319-67558-9_28. ISBN: 9783319675589. http://dx.doi.org/10.1007/978-3-319-67558-9_28.
- [6] Walsh, C., Tafforeau, P., Wagner, W., Jafree, D., Bellier, A., Werlein, C., Kühnel, M., Boller, E., Walker-Samuel, S., Robertus, J., et al.: Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nature Methods* 18(12), 153 – 1541 (2021)
- [7] Nicholson, C.: Diffusion and related transport mechanisms in brain tissue. *Reports on Progress in Physics* 64(7), 815 (2001)
- [8] Nikolov, S., Blackwell, S., Zverovitch, A., Mendes, R., Livne, M., De Fauw, J., Patel, Y., Meyer, C., Askham, H., Romera-Paredes, B., Kelly, C., Karthikesalingam, A., Chu, C., Carnell, D., Boon, C., D’Souza, D., Moinuddin, S.A., Garie, B., McQuinlan, Y., Ireland, S., Hampton, K., Fuller, K., Montgomery, H., Rees, G., Suleyman, M., Back, T., Hughes, C., Ledsam, J.R., Ronneberger, O. (2021) Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. *arXiv preprint*, arXiv:1809.04430, 2021.