# Supplementary material

## Feature distance of image patches for tracklet merging

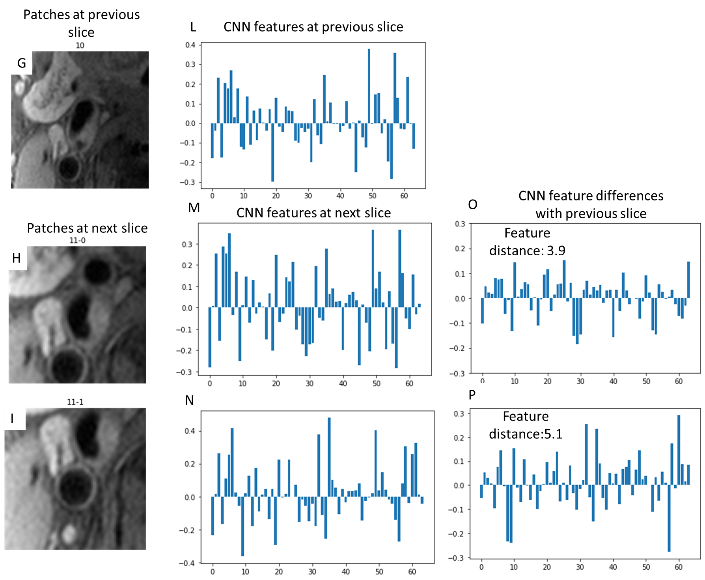


Fig. 1 CNN feature and feature distances for patches from Figure 3 in the main manuscript. L, M, N are the 64-dimensional features for the patches shown in G, H, I. O and P are the feature distance between L vs M and L vs N. Tracklets are merged with smaller feature distance. In this case, L is merged with M.

## Polar conversion

The polar image patches with the size of are converted from cartesian image patches , using the polar transformation relation of

The inverse polar conversion for patches is defined as , where

## Upper limit of polar regression

Upper limit of vessel wall segmentation through polar regression is shown in Table 1. We observe that once the number of directions is beyond 64, there is little difference in terms of DSC with ground truth. The high DSC (>0.96) indicates the feasibility of using polar regression on vessel wall segmentation without much deterioration of labels. So the number of 256 directions is used, which is the same as the largest value in axis to maintain the square shape of polar patches. Compared with the upper limit, our best polar regression model still has much room for improvement, so the loss of accuracy in the polar conversion is not our top concern for now.

Table 1 Upper limit of vessel wall segmentation through polar regression

|  |  |  |  |
| --- | --- | --- | --- |
| Number of directions | Mean DSC | Percent of DSC<0.93 | Percent of DSC<0.95 |
| 32 | 0.9597 | 1.54% | 20.20% |
| 64 | 0.9623 | 0.44% | 14.18% |
| 128 | **0.9630** | **0.24%** | **12.91%** |
| 256 | 0.9626 | 0.50% | 14.06% |
| 512 | 0.9618 | 0.56% | 15.36% |

## Evaluation metrics

Dice similarity coefficient (DSC) [1] is defined as

where A is the ground truth result and B is the segmentation result. DSC ranges from 0 (no overlap) to 1 (identical results). DSC > 0.7 indicates excellent agreement [1].

Degree of similarity (DoS) [2] is defined as

N is the number of paired sample points on the predicted contour and the ground truth contour, = 100 is used in literature and in this study. Each pair of points on the contours have a distance of . is a distance threshold to allow the errors of contour prediction, which is chosen as the same as the imaging resolution as the common practice.

***Dataset detailed property***

TABLE

Imaging parameters and properties for datasets used in this study

|  |  |  |  |
| --- | --- | --- | --- |
|  | CARE-II | Kowa | |
| Scanned artery region | Carotid artery | | | |
| Sequence | Fat suppressed T1 weighted Turbo Spin Echo | | | |
| Blood suppression | Quadruple inversion recovery | | | |
| Number of subjects | 954 | | 203 |
| Number of labeled slices | 27397 | | 5194 |
| TR (ms) | 800 | | 800 |
| TE (ms) | 10 | | 10 |
| In-plane Resolution (mm) | 0.57\*0.57 | | 0.63\*0.63 |
| Spacing between slices (mm) | 2 | | 2 |
| Field of view (mm\*mm) | 160\*160 | | 160\*160 |

## Rotated patch with different gaps

Parameter used in combing patch rotation for prediction is selected based on the tradeoff between calculation time, accuracy and uncertainty scores. Lower means more overlapping rotated patches are used in segmentation, which costs more time but could generate a smoothed probability map and/or contours by averaging from different patches (glitches from a few rotated patches might be subdued). In addition, with enough samples of predictions from multiple patches, the distribution of boundary can be better reflected. On the other extreme, predicting without overlap () cannot calculate boundary consistency at all, and segmentation from a single patch may lead to rough probability map and/or boundaries. Different values are selected (from 1 to ) and evaluated from the validation set of the carotid dataset, then a best tradeoff value is used to apply to the test set. The set of DSC (vessel wall, lumen, and outer wall), DoS (lumen and wall), correlation coefficients of uncertainty scores with DSC (CC-Lumen, CC-Wall), and the calculation time is shown in Table 3.



Fig. Contour consistency with DSC from the Polar-Reg CNN architecture.



Fig. Decreased DSC and uncertainty scores (Cst\_l/Cst\_w: boundary consistency for lumen/outer wall) with increased level of noises from the Polar-Reg CNN architecture.

## Segmentation uncertainty scores

The scatter plots between lumen and wall consistency scores with DSCVW for the Polar-Reg CNN architecture is shown in Fig. 1. Generally, lower boundary consistencies indicate poor segmentations having low DSC compared with human labeled vessel wall. Lumen consistency score has the correlation coefficient of 0.132 (p<1e-5), and wall consistency score has the correlation coefficient of 0.047 (p=0.045).

To further prove the sensitivity of consistency scores when the vessel wall image is of lower quality, we added different levels of random noises following the Rayleigh distribution with mode of 1 to the existing vessel wall images. The noise added image where follows . DSCVW and mean of each of three uncertainty scores are evaluated from different noise levels at [0,0.2,0.4,0.6,0.8,1]. The line plot of results is shown in Fig. 2.

Table 3 Comparison between gaps of rotated patch from the validation set

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gap | DSCVW | DSCInner | DSCOuter | DoSLumen | DoSWall | CC-Lumen | CC-Wall | Time (s) |
| 1 | 0.824 | **0.945** | 0.949 | 0.856 | 0.776 | 0.449 | 0.350 | 3.06 |
| 2 | 0.824 | 0.945 | 0.949 | 0.856 | 0.776 | 0.448 | 0.350 | 1.57 |
| 5 | **0.824** | 0.945 | **0.950** | 0.856 | 0.776 | 0.448 | 0.346 | 0.68 |
| 10 | 0.824 | 0.945 | 0.950 | 0.857 | 0.776 | 0.449 | 0.343 | 0.37 |
| 20 | 0.823 | 0.945 | 0.949 | **0.857** | 0.774 | 0.436 | 0.334 | 0.22 |
| 32 | 0.823 | 0.944 | 0.949 | 0.855 | 0.774 | **0.469** | **0.384** | 0.17 |
| 64 | 0.823 | 0.944 | 0.949 | 0.854 | 0.776 | 0.411 | 0.318 | 0.12 |
| 128 | 0.823 | 0.944 | 0.949 | 0.853 | **0.778** | 0.356 | 0.206 | 0.09 |
| 256 | 0.820 | 0.943 | 0.948 | 0.851 | 0.772 | N/A | N/A | **0.08** |

## Reference

[1] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[2] E. Angelie, E. R. Oost, D. Hendriksen, B. P. F. Lelieveldt, R. J. Van der Geest, and J. H. C. Reiber, “Automated Contour Detection in Cardiac MRI Using Active Appearance Models,” *Invest. Radiol.*, vol. 42, no. 10, pp. 697–703, 2007.