1. Supplementary Material

1.1. Revolving Attention Block (RAB)

The revolving attention block is a three-branched feature extraction and attention network. To leverage *what* & *where to focus*

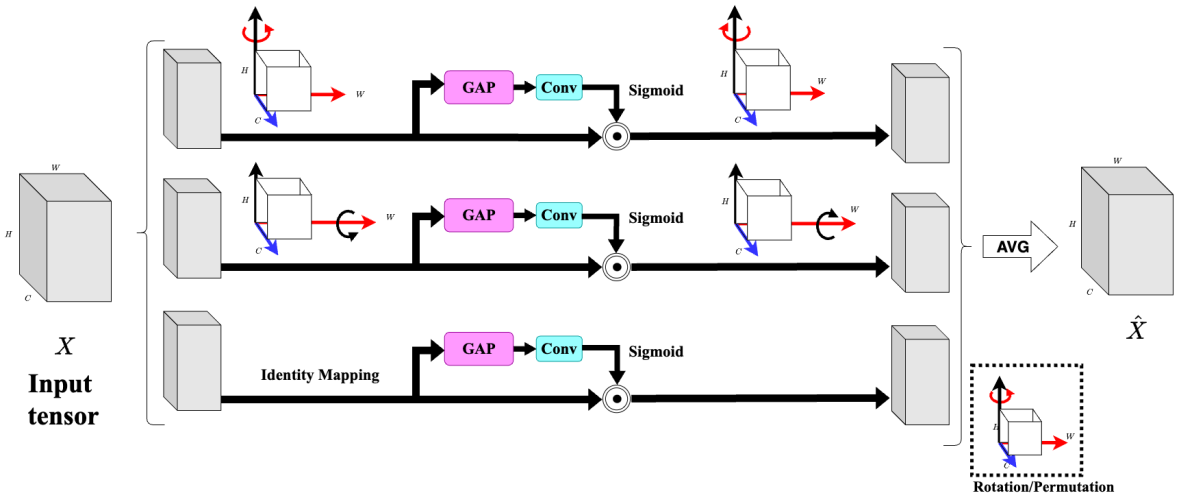

Figure 1: The revolving attention block (RAB) with triplet attention mechanism, attends the input feature channels by rotating them into 3 specific orders shown as 3 branches of the architecture.

in the spatial and channel domains, it processes the input X in height, width, and channel dimension. The RAB block avoids the loss of spatial information, which was evident in traditional cross-dimension interaction by computing a singular weight for each channel [1]. The RAB attention block receives the permuted input tensors with order $(C \times H \times W), (H \times C \times W)$ and $(H \times W \times C)$ in three branches, as shown in Fig. 1. The attention inside the RAB block is executed using global and average pooling block which reduces the zeroth dimension of the input X to two, resulting in tensors of the order$(2 \times H \times W), (2 \times C \times W)$ and $(2 \times W \times C)$ for the respective branch. The pooling operation is shown as;

$$X_{pool}(X) = \{MaxPool(X[0]), AvgPool(X[0])\} \tag{1}$$

The resulting tensors are then processed through a convolution with kernel size k, an instance-norm, and a sigmoid activation function. The produced outputs termed as attention weights of order $(1\times H\times W)$, $(1\times C\times W)$, and $(1\times W\times C)$, are then applied to respective branches highlighting the important features in each branch. The resulting features are permuted to their original geometry shown in Fig. 1, and the final output $\hat{X}$ is the averaged output from all branches.

1.2. Effect of loss function coefficients on training.

In this paragraph, we show the effect of different loss coefficients in the loss optimization equation in the main section. We show the variation of DICE score while validation runs with different values of λ_2, λ_3, and λ_4 as listed in Fig. 2, 3, and 4, respectively. The loss controlled by λ_2 is the cycle loss, which is an integration of three losses as detailed in the main section. The cycle loss helps the model to converge faster and punish the segmentation generation with *cldice* loss. Whereas λ_3 controls the contrastive loss, helping the architecture to generate vessel segmentations overall.

1.3. Fractal Images used as pretext task for Self-Supervised Learning

We use these fractals for the self-supervised training, these fractals provided the structural similarity with the coronary angiograms while optimising the model's weights. Some example fractals are shown in Fig. 5.

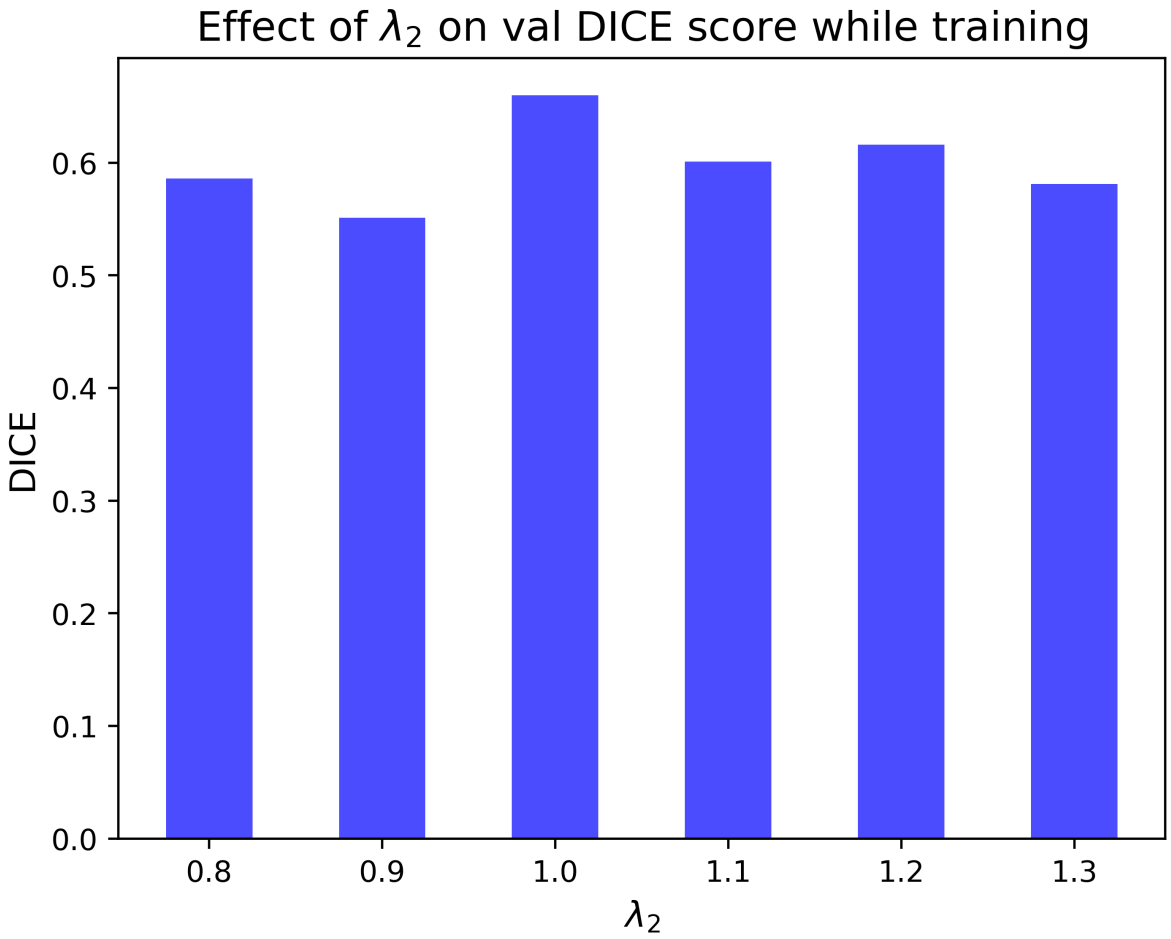

Figure 2: The validation DICE values with varying λ_2.

1.4. Inference on local hospital data

The local data collected from MCH Calicut, India, is anonymized with personal details such as names and addresses of the patients. We collected 200 angiographic series with a mix of present stenosis. Each series of coronary studies has been taken from multiple orientations such as cranial and caudal views of the left and right coronary arteries from right and left anterior oblique body orientation. The selected subjects span from 20-70 years of age and were randomly selected among gender. Each series is validated by experienced surgeons, and the stills selected from the series have been of appropriate quality to be processed. We show some samples from the dataset being segmented in Fig. 6. Currently, the process of data labelling and legal processes for releasing the dataset are underway.

References

[1] D. Misra, T. Nalamada, A. U. Arasanipalai, Q. Hou, Rotate to attend: Convolutional triplet attention module (2020). arXiv:2010.03045. URL https://arxiv.org/abs/2010.03045

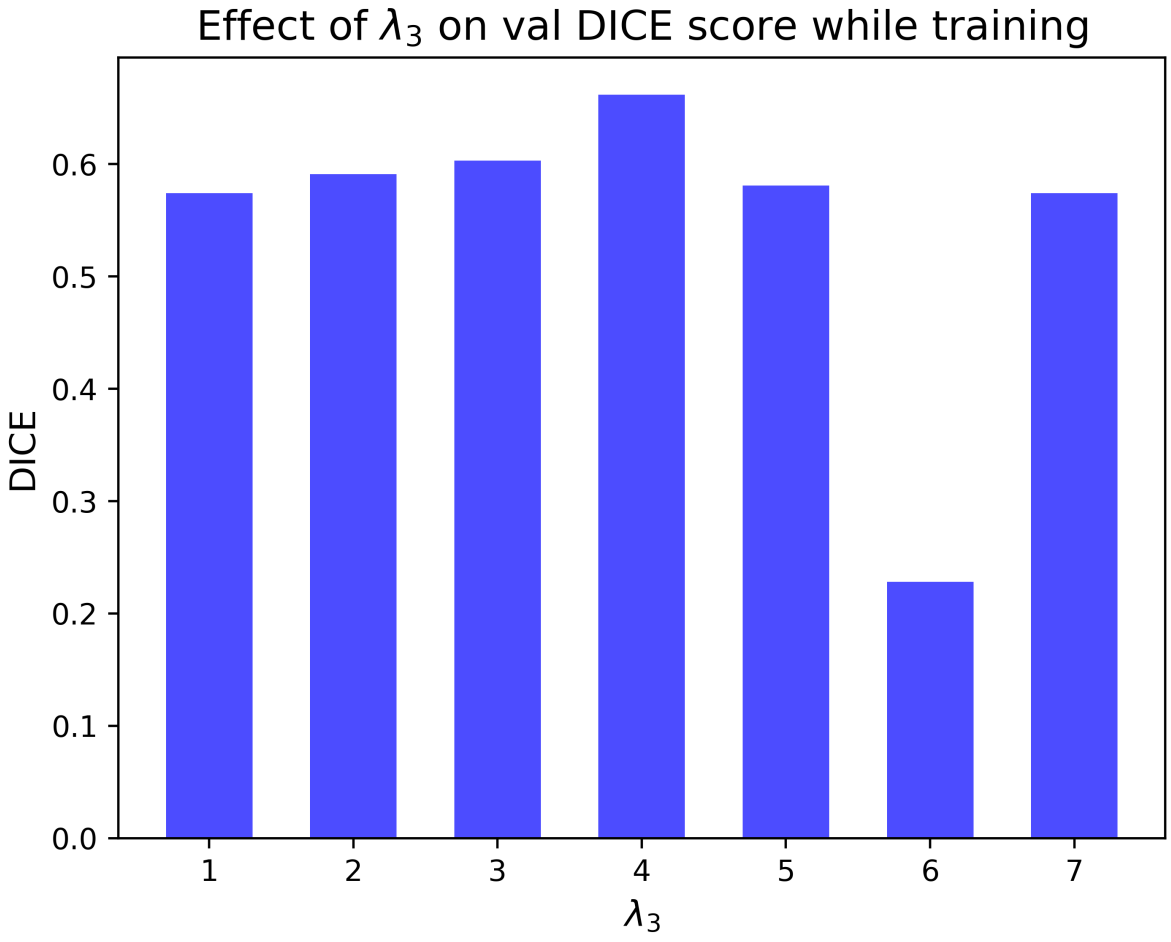

Figure 3: The validation DICE values with varying $\lambda_3(e^{-4})$.

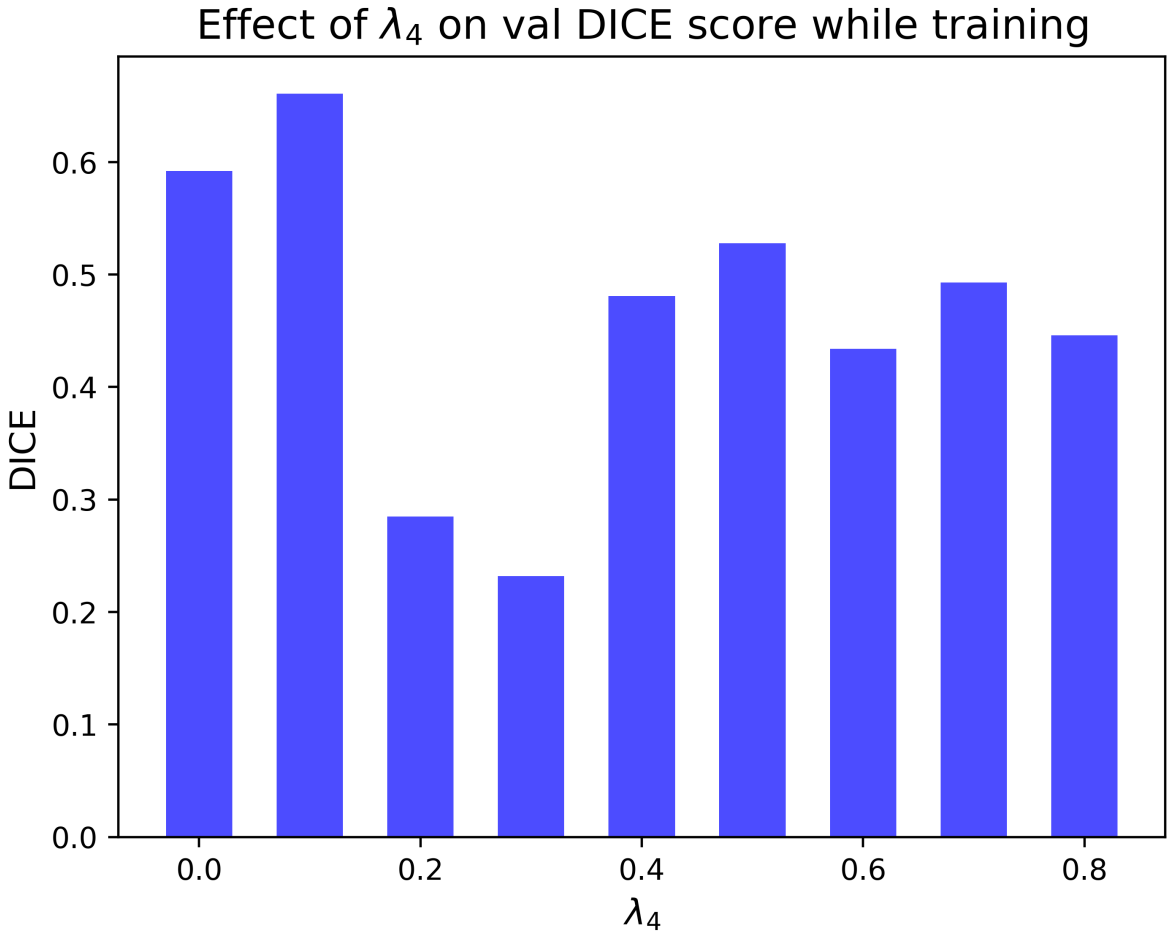

Figure 4: Validation DICE score values while training with varying λ_4 coefficient.

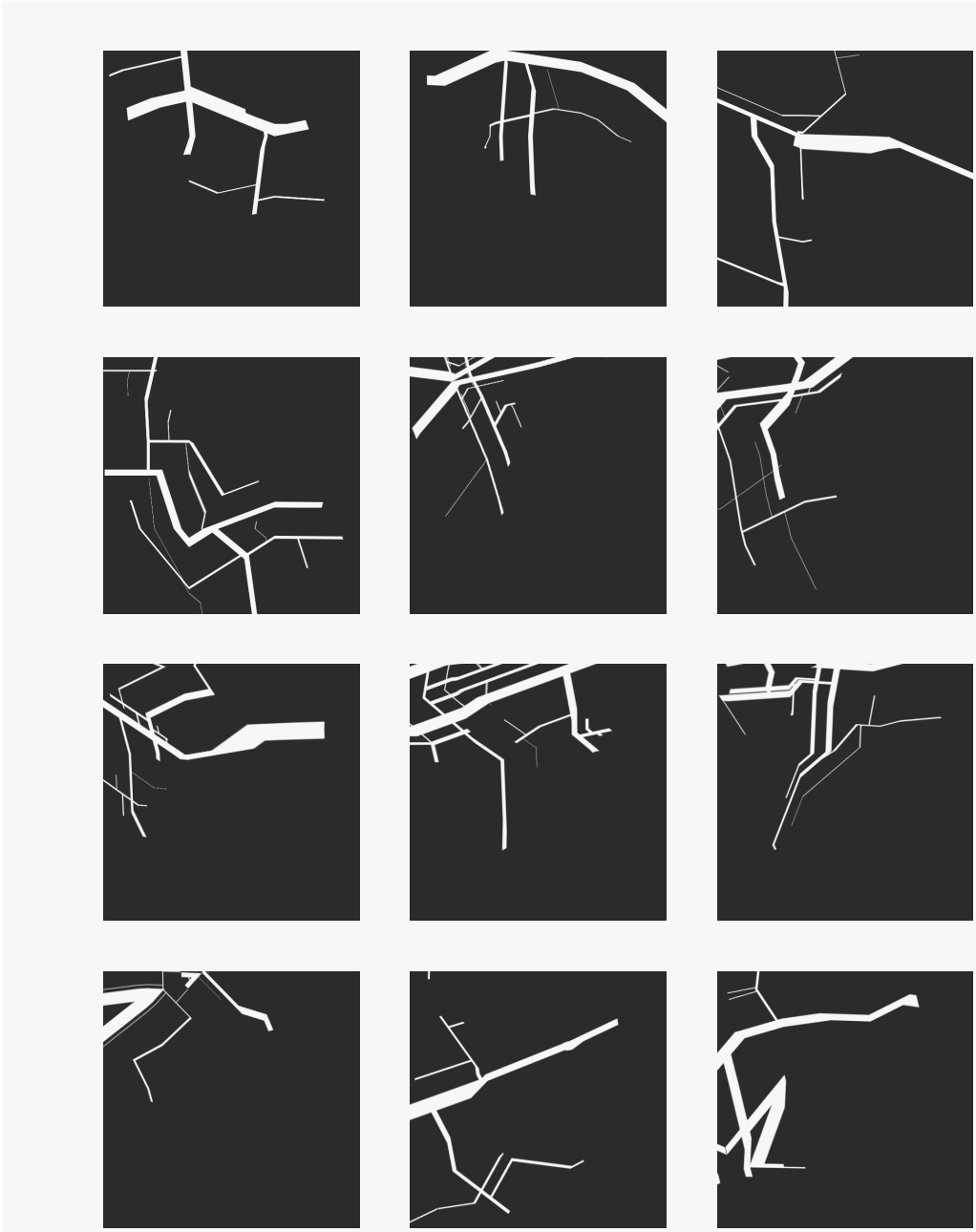

Figure 5: Sample fractals utilised to train the segmentation model.

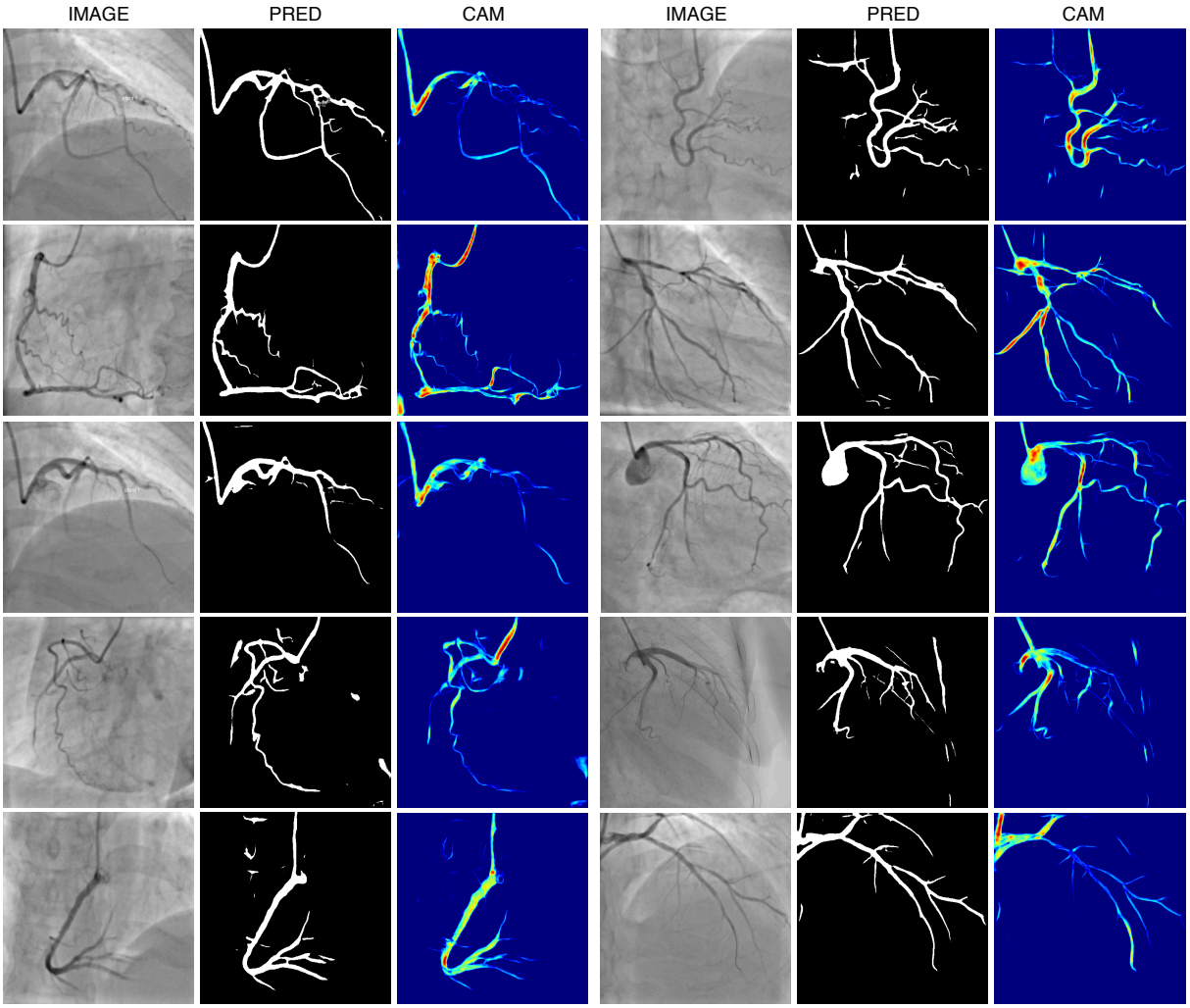

Figure 6: Segmentation results on local data from MCH, showing promising output segmentation maps.