

Real-time coronary artery segmentation in CAG images: A semi-supervised deep learning strategy

Chih-Kuo Lee^{a,c,d}, Jhen-Wei Hong^b, Chia-Ling Wu^b, Jia-Ming Hou^b, Yen-An Lin^b, Kuan-Chih Huang^a, Po-Hsuan Tseng^{b,*}

^a Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, No. 25, Lane 442, Section 1, Jingguo Rd, North District, Hsinchu City 300, Taiwan

^b Department of Electronic Engineering, National Taipei University of Technology, No. 1, Sec. 3, Chunghsiao E. Rd., Taipei City 10608, Taiwan

^c Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, No.1, Chang-Te St., Taipei 100, Taiwan

^d Department of Internal Medicine, College of Medicine, National Taiwan University, No.1, Jen Ai road section 1, Taipei 100, Taiwan



ARTICLE INFO

Keywords:

Semi-supervised learning
Consistency regularization
Pseudo-labeling
Semantic segmentation
Coronary angiography
RandAugment

ABSTRACT

Background: When treating patients with coronary artery disease and concurrent renal concerns, we often encounter a conundrum: how to achieve a clearer view of vascular details while minimizing the contrast and radiation doses during percutaneous coronary intervention (PCI). Our goal is to use deep learning (DL) to create a real-time roadmap for guiding PCI. To this end, segmentation, a critical first step, paves the way for detailed vascular analysis. Unlike traditional supervised learning, which demands extensive labeling time and manpower, our strategy leans toward semi-supervised learning. This method not only economizes on labeling efforts but also aims at reducing contrast and radiation exposure.

Methods and results: CAG data sourced from eight tertiary centers in Taiwan, comprising 500 labeled and 8952 unlabeled images. Employing 400 labels for training and reserving 100 for validation, we built a U-Net based network within a teacher-student architecture. The initial teacher model was updated with 8952 unlabeled images inputted, employing a quality control strategy involving consistency regularization and RandAugment. The optimized teacher model produced pseudo-labels for label expansion, which were then utilized to train the final student model. We attained an average dice similarity coefficient of 0.9003 for segmentation, outperforming supervised learning methods with the same label count. Even with only 5 % labels for semi-supervised training, the results surpassed a supervised method with 100 % labels inputted. This semi-supervised approach's advantage extends beyond single-frame prediction, yielding consistently superior results in continuous angiography films.

Conclusions: High labeling cost hinders DL training. Semi-supervised learning, quality control, and pseudo-label expansion can overcome this. DL-assisted segmentation potentially provides a real-time PCI roadmap and further diminishes radiation and contrast doses.

1. Introduction

Coronary artery disease is the most prevalent cardiovascular disease worldwide and is caused by coronary artery stenosis (i.e., narrowing or occlusion of the arteries). Coronary angiography (CAG) is the benchmark imaging modality in assessing the severity of coronary artery stenosis via continuous X-ray images. High-quality CAG images play an

important role in guiding clinical decisions when cardiologists perform percutaneous coronary intervention for coronary artery disease and acute coronary syndrome, such as stent sizing and the strategy choice between stent deployment and balloon angioplasty only. CAG images not only assist in visualizing the vascular anatomy but also serve as the foundation for subsequent analyses. Therefore, precisely extracting the coronary artery structure from these images is crucial, as it yields

Abbreviations: DSC, dice similarity coefficient; S3, semi-supervised semantic segmentation; FCNs, fully convolutional networks; EMA, exponential moving average; MSE, mean square error; UDA, unsupervised data augmentation; VAT, virtual adversarial training; MPL, meta pseudo label; CPS, cross-pseudo supervision; GDL, generalized dice loss.

* Corresponding author at: 1, Sec. 3, Zhong-Xiao East Road, Taipei 10608, Taiwan.

E-mail address: phtseng@ntut.edu.tw (P.-H. Tseng).

<https://doi.org/10.1016/j.artmed.2024.102888>

Received 28 September 2023; Received in revised form 24 April 2024; Accepted 30 April 2024

Available online 9 May 2024

0933-3657/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

quantitative information vital for tailoring treatment strategies to effectively address patient-specific conditions.

A CAG image is a projection of the three-dimensional (3D) coronary artery anatomy containing the flow status of contrast media onto a two-dimensional (2D) image; despite the inherent dimensional reduction from 3D to 2D in CAG images, which results in a loss of some volumetric information, a detailed and comprehensive analysis of 2D CAG images through deep learning might partially compensate for this loss. The essence of our approach lies in extracting maximal information from the 2D data available, leveraging advanced image analysis techniques to discern patterns and features that are not visually apparent. This method of in-depth analysis aims to bridge the gap caused by the absence of the third dimension, and there is existing medical literature supporting the premise that such detailed analyses can outperform the visual judgment of cardiologists in certain clinical settings [1]. This is particularly beneficial in cases where patients with poor renal function cannot tolerate large amount contrast medium and multi-angle angiography. Conventional methods such as the Hessian filter with Otsu thresholding [2] have been applied for accurately detecting the artery using the local pixel neighborhood. Recently, supervised semantic segmentation using deep learning has been applied to this problem. Several deep-learning-based vessel segmentation methods were proposed, such as attention-based network [3–6], or unrolling network [7]. Because stenotic lesions are easily found in the main coronary arteries (main vessels of LAD, LCX and RCA), numerous studies [8,9] have focused on the segmentation of the three main coronary arteries only. Various semantic segmentation models, including UNet and its variants (used to replace the backbone network), have been applied [9]. However, all branches related to these three main coronary arteries are also crucial for comprehensive physiological quantitative analysis, such as that required for coronary blood flow analysis or dynamic roadmap. Full-branch vessel labels were considered in some studies [10,11]. The main challenge of full branch segmentation comes from the small vessel segmentation and topology preservation, partly because of the diversity and complexity of distal-branch anatomy for the distinguishment with background interference. Moreover, labeling is time-consuming, and a consistent labeling system has not been established. Compared to the abundant label numbers in main coronary arteries (3302 labeled images [9]), the number of full-branch labels is 148 [10] and 60 [11]. Therefore, this study introduces a comprehensive segmentation design targeting full-branch vessels, a strategy poised to enhance further quantitative analyses. Recognizing the constraints imposed by the limited resolution and sample size of the presently accessible public datasets—namely DCA1 [12], encompassing 134 CAG images with 300×300 pixel resolution, and CHUAC [13], which includes 30 CAG images at a 189×189 pixel resolution — our objective is to cultivate a repository of high-quality images.

Given that CAG films are a series of images with a higher number of input images, films need to be taken from 4 to 6 different angles for every patient, with each film containing at least 60 frames for the observation. Although sufficient sample collection is possible, the expensive and labor-intensive pixel-wise labeling that requires medical domain knowledge hinders the construction of adequate labels for supervised learning. Semi-supervised learning seeks to enhance learning from a limited labeled dataset by incorporating insights drawn from a larger unlabeled dataset. The medical segmentation tasks generally adopt consistency regularization, Mean Teacher (MT) [14], achieves quality control by adopting the unlabeled data to learn the generalized feature and remove the outlier. In addition to introducing the Gaussian noise of MT, unsupervised data augmentation (UDA) [15] implements RandAugment as the unlabeled data enhancement method and as its parameter range selection to impose noise on the input data. Given that further combining consistency regularization with pseudo-labels and knowledge distillation leverage unlabeled data obtained from the teacher-student network, Meta Pseudo Label (MPL) [16] utilizes UDA combined with the pseudo-labels generated in a teacher-student model

training with the addition of meta-learning. Further, virtual adversarial training (VAT) [17] considers the regularization in terms of models' divergence-based distribution against the virtual adversarial perturbation. The concept of self-training [18] is extended to two networks that output the pseudo-labels and serve as the supervision by using the cross-entropy loss constraints for each other in cross-pseudo supervision (CPS) [19]. A semi-supervised semantic segmentation using self-training (ST) [20] with strong data augmentations on unlabeled images reduces the overfitting of noisy labels between the teacher and student models. By estimating image-level uncertainty through holistic prediction stability, the selective retraining of ST++ [20] further prioritizes reliable unlabeled images from pseudo-label without the need for manual threshold selection.

With the design of semi-supervised semantic segmentation, we aim to systematically answer the question of achieving optimal performance with reasonable cost, that is, labeling cost-performance tradeoffs under the real-time computation constraint. To this end, we are pioneering a semi-supervised semantic segmentation (S3) approach. This innovative methodology not only leverages the restricted set of labeled data at our disposal but also harnesses the substantial volume of unlabeled data for augmented training. Our aspiration is to realize peak performance while maintaining a rational expenditure, effectively balancing the cost-performance dynamics of the labeling process.

The contributions of our work are as follows.

- We have established a unique and extensive CAG image database from eight tertiary centers in Taiwan, which is unprecedented worldwide, comprising 500 labeled and 8952 unlabeled images. The labeling process was carried out by experienced technicians from our central laboratory. These labels then underwent rigorous review during consensus meetings led by a team of cardiologists and radiologists. By using a U-Net based network within a teacher-student architecture and employing only 400 labels for training (with 100 reserved for validation), we implemented a semi-supervised learning framework that effectively overcomes the limitations associated with the scarcity of labeled medical data and the constraints of annotation resources.
- Our innovative approach, complemented by a quality control strategy that includes consistency regularization and RandAugment. With 9452 CAG images, we trained the model by labeling only 5.3 % of the entire data and fully utilizing the rest unlabeled images to attain an average dice similarity coefficient of 0.9003 for segmentation. This performance significantly surpasses that of supervised learning methods using an equivalent count of labels.
- Emphasizing real-time processing, our approach analyzes CAG images in sync with the heart's beating rate. The advantage of this semi-supervised method extends beyond predictions for single CAG images, yielding consistently superior results across intra-surgery continuous CAG films. This approach not only represents a significant step forward but also shows high potential for further enhancing patient safety by minimizing exposure to contrast agents and radiation. Additionally, it provides solutions that bridge the gap between computational methods and medical applications.

2. Method

2.1. S3: semi-supervised semantic segmentation for cag images

S3 is a deep segmentation method for accurate vessel segmentation of CAG with fewer labeled images. T and S represent the considered teacher and student models, respectively, associated with their respective parameters θ and θ_S , as shown in Fig. 1. S3 is a hybrid method involving consistency regularization and self-training, with the uncertainty of unlabeled data being considered by the designed two-way masking and threshold.

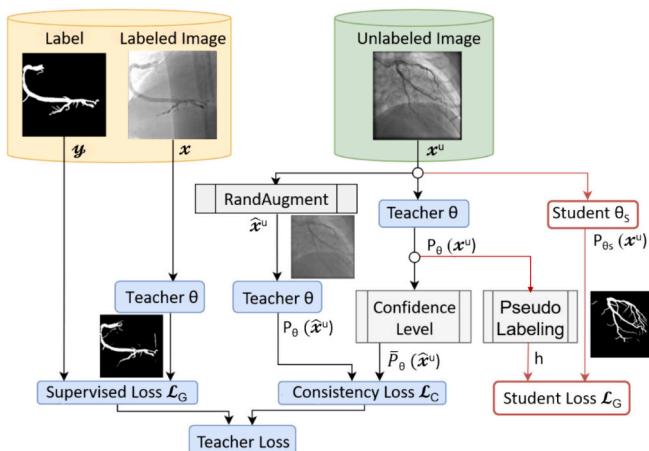


Fig. 1. Proposed S3 method for coronary artery segmentation in CAG images.

2.2. Initialization of teacher model

In the proposed S3, the first step is to establish the initial teacher model using a UNet-based backbone and input labeled data. As a semantic segmentation model, UNet generally comprises an encoder (downsampling) and a decoder (upsampling). To avail the advantages of UNet, our teacher model in the semi-supervised learning process replaces the backbone with SOTA image classification models like EfficientNet [21], ResNet [22], DenseNet [23], and ResNeSt [24]. EfficientNet [21] uniformly scales the network width, depth, and resolution with a set of fixed scaling coefficients for better performance. ResNet [22] utilizes the skip connection with a convolutional neural network; its variants include DenseNet [23] (dense connections of shortcuts among layers) and ResNeSt [24]. The former has dense connections of shortcuts among layers; the latter further applies the attention mechanism to the channels of different network branches to capture the representations and cross-feature interactions. ResNeSt is a SOTA method in terms of accuracy-latency tradeoffs, and is effective as the backbone of the object detection problem.

The segmentation prediction of the UNet-based backbone network is denoted as $P_\theta(x)$, where θ represents the network parameters, (x, y) represents the input batch of images and associated labels, and (x_i, y_i) represents the i -th pixels of (x, y) . The input and output sizes of this network are identical, and N is the total number of pixels in the input image. In coronary artery segmentation from the CAG image, the binary segmentation considers class $y_i = 1$ as a coronary artery pixel i (colored white) and class $y_i = 0$ as a non-vessel background pixel i (colored black). Because the CAG images are black and white for a binary-class segmentation, the sigmoid is adopted as the activation function in the last layer to map the predicted output from the black and white colors to the values of 0 or 1. To consider the imbalance between the pixel numbers that represent the coronary arteries and those that do not, we adopted the generalized dice loss (GDL) [25]. $L_G(p, r)$ denotes the supervised loss of the teacher model between the predicted target and reference distributions p and r , it is expressed as

$$L_G(p, r) = 1 - 2 \frac{w_1 \sum_{i=1}^N p_i \cdot r_i + w_0 \sum_{i=1}^N (1 - p_i) \cdot (1 - r_i)}{w_1 \sum_{i=1}^N (p_i + r_i) + w_0 \sum_{i=1}^N (2 - p_i - r_i)} \quad (1)$$

where p_i and r_i represent the i -th pixels of the prediction and the label, respectively. The formulation is represented by treating the binary segmentation as two classes, 1 and 0, and mapping them as (p, r) and $(1 - p, 1 - r)$, respectively. The weights of classes 1 and 0 depend on the label distribution as $w_1 = 1 / (\sum_{i=1}^N r_i)^2$ and $w_0 = 1 / (N - \sum_{i=1}^N r_i)^2$, respectively. If there are multiple instances of p and r , $L_G(p, r)$ is the average of the instances in a batch. The clamping term can be further

used in the denominator to avoid dividing the value by zero. In the considered vessel segmentation problem, we observed that the UNet-based backbone combined with the GDL [25] has superior DSC for medical images.

2.3. Consistency regularization for CAG segmentation

We added unlabeled data in the teacher model and set up a quality-improvement mechanism—consistency regularization. Besides the labeled samples utilizing GDL in (1), the unlabeled images are provided as two inputs to the network: (i) the input using RandAugment and (ii) the original input without any processing. Next, the consistency loss is determined by the similarity between the predictions from the original unlabeled input x^u and that after RandAugment \hat{x}^u is applied. We used only the high-confidence pixels agreed by the predictions of different types of RandAugment to update the teacher model, which lowers the prediction uncertainty. Subsequently, the consistency loss is determined by calculating the pixel-wise MSE loss with confidence masking between the prediction from the original unlabeled sample and that from its augmented version as

$$L_C(\theta) = \sum_{i=1}^N m_i \cdot \|P_\theta(x_i^u) - P_\theta(\hat{x}_i^u)\|^2. \quad (2)$$

The semi-supervised network is trained by using the loss function $L_G(y, P_\theta(x)) + \omega \cdot L_C(\theta)$, where ω is the weight between the two losses. We designed two-way confidence-based masking by modifying the confidence masking in the classification problem [15] for the predicted result. The confidence-based masking of the i -th pixel m_i in (2) is defined as

$$m_i = I(\{P_\theta(x_i^u) \geq \tau\} \cup \{P_\theta(x_i^u) \leq 1 - \tau\}), \quad (3)$$

where $I(\cdot)$ is the indicator function for the confidence level, i.e., values smaller than a threshold τ or larger than $1 - \tau$ have high confidence level, at which the pixel is correctly predicted.

We applied RandAugment by adding noise and variation to the image to ensure that the model could learn the vessel perfectly, even under different background settings, thus making the learning process more robust. To add a suitable random disturbance to the input image, we tested RandAugment on randomly selected data augmentation methods and their parameters for coronary artery segmentation to utilize unlabeled samples. A type of RandAugment is geometric data augmentation, such as rotational, horizontal, and vertical displacement. It is efficient in image classification and supervised image segmentation problems. In the latter case, the input and its label can be augmented with the same geometric adjustment to increase the data variety. However, the semi-supervised vessel segmentation task relies on the single-pixel binary classification in the overall contour. Assume that after the RandAugment input, the vessel pixel moves to a different place while the original input uses geometric augmentation. This violates the requirement that the information on the two sides of consistency regularization must have pixel-wise consistency in semi-supervised image segmentation. Therefore, the RandAugment data in this study do not include geometric data enhancement. Fig. 2 depicts the various RandAugment types considered in this study. We verified in the experiment that blur and contour are considered suitable for the CAG segmentation, and the detailed results can be found in Supplemental Table S1.

2.4. Self-training with teacher-student networks

We used this teacher model to generate high-quality pseudo-labels, expand the label bank, and train a complete student model utilized for online prediction. Although different networks can be employed in the teacher-student model, we adopted the same networks. The student model was set up with greater complexity than the teacher model, e.g., noisy student method [26]; if the teacher model was ResNeSt101 or EfficientNetB5, the student model was EfficientNetB6.

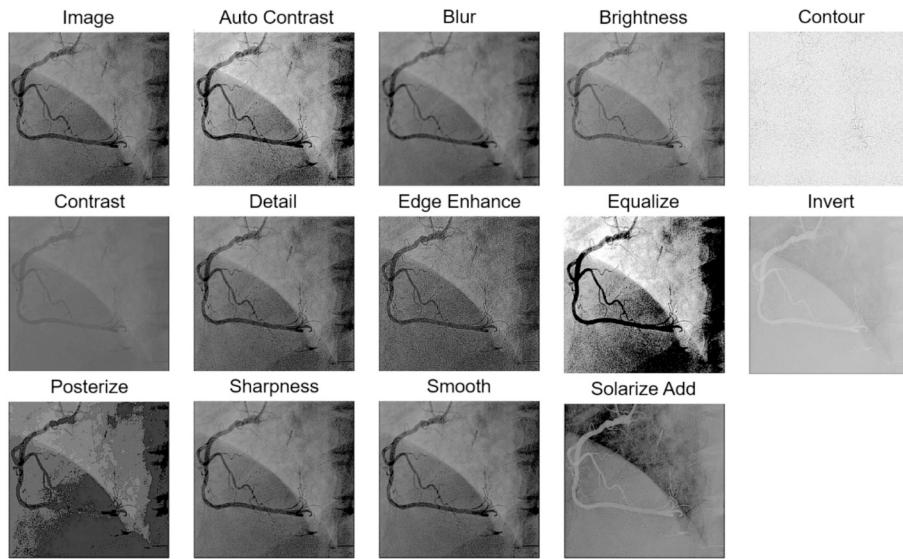


Fig. 2. Random augmentation methods for consistency regularization used in this study. The contour plot could be observed by enlarging the figure.

By further adopting data distillation, the prediction of unlabeled data from the teacher network serves as the pseudo-label to train the student network. Unlike the soft-label strategy used in confidence masking, the hard pseudo-label is obtained as

$$h_i = \begin{cases} 1, & P_\theta(x_i^u) \geq \tau_h \\ 0, & P_\theta(x_i^u) < 1 - \tau_h \end{cases} \quad (4)$$

Given the set of pixel indices for the input images, $N = \{1\dots N\}$, and the threshold τ_h , we denote the set of indices for the vessel pixel class 1 as $C_1 = \{i \in N : h_i = 1\}$ and that of the background class 0 as $C_0 = \{I \in N : h_i = 0\}$. Thus, the GDL between the pseudo-label h and output of the unlabeled image p can be determined by

$$L_G \left(h, p \right) = 1 - 2 \frac{w_1 \sum_{i \in C_1} h_i \cdot p_i + w_0 \sum_{i \in C_0} (1 - h_i) \cdot (1 - p_i)}{w_1 \sum_{i \in C_1} (h_i + p_i) + w_0 \sum_{i \in C_0} (2 - h_i - p_i)}, \quad (5)$$

where $w_1 = 1/|C_1|^2$, $w_0 = 1/|C_0|^2$; $|C_1|$ and $|C_0|$ are the number of elements in the set. In the student loss, we represent the GDL between the pseudo-label and the student network's prediction of the unlabeled image as $L_G(h, P_{\theta_s}(x^u))$. Combined with the use of consistency loss and RandAugment as the total loss as $L_G + \alpha L_C$, where the consistency weight is defined as α , we avoid the model suffering from confirmation bias on generating pseudo labels.

3. Experiment settings and results

Our model shows a smooth segmentation result for the main coronary artery and the full-branch vessel in continuous X-ray images of CAG. This was achieved by using unlabeled frames (different from the keyframe) without additional labeling effort for every frame of the continuous CAG films. By eliminating unnecessary tasks, the application of our model to CAG films could provide a dynamic roadmap in performing percutaneous coronary intervention and reducing the contrast medium dosage.

3.1. Experimental setup

1) *Clinical Dataset:* The data used here were derived from the Medical Imaging Project database of the Ministry of Science and Technology, Taiwan, established by eight medical centers, which were involved in collecting CAG images, including the National Taiwan University

Hospital and Chang Gung Memorial Hospital. Totally, 213 patients were enrolled. The CAG X-ray imaging was acquired from multiple angles to facilitate viewing the structures of coronary arteries. These results are stored in the Digital Imaging and Communication in Medicine (DICOM) format with a frame rate of 15 fps, the resolution of each frame being 512×512 .

- 2) In the experimental setup, we train the proposed S3 network using NVIDIA-DGX with 4 V100. The epoch is chosen as 200. The SGD optimizer is adopted with the learning rate of the teacher's network of 0.08 and that of the student's network of 0.1. The consistency weight α is chosen as 1, in the study as we found its smooth convergence as the epoch increases.
- 3) *Labeled and Unlabeled Image Acquisition:* The vessel-labeling process included the main arteries and full-branch vessels with diameters larger than 2.0 mm. The labeling software supported the forward and backward scrolling among the frames of the CAG films, allowing the image analysts to eliminate any feature not associated with the movement of vessels expected in a regular heartbeat pattern (i.e., non-vascular artifacts) through comparison between adjacent frames. To reach a consensus on non-vascular artifacts, the labels were examined by at least three cardiologists or radiologists.

We labeled 500 frames by selecting the end-diastolic and -systolic frames when the contrast medium completely filled the lumen of coronary arteries. For example, there were 107 frames in one RCA-RAO Cranial angiography film stored in a DICOM file, and the diastolic (number 55) and systolic frames (number 62) were selected as the keyframes to be labeled. Five hundred labeled images had 220 RCA-RAO Cranial, 220 LCA-RAO Caudal, and 60 LCA-RAO Cranial angiographies. To fully utilize the unlabeled frames in the DICOM files, we also selected the unlabeled data from the remaining frames. If the DICOM file included a labeled frame, the preceding and succeeding frames were incorporated into the unlabeled dataset. Otherwise, the frame with the fullest contrast medium in the lumen of the coronary artery and 10 preceding and succeeding frames were regarded as unlabeled data. Further, the 8952 unlabeled images included 5211 LCA-RAO Caudal, 3243 RCA-RAO Cranial, and 498 LCA-RAO Cranial angiographies.

- 4) *Segmentation Performance Evaluation and Visualization:* In the vessel segmentation task, we used the model to predict images pixel by pixel and classify the vessel pixels as positive and non-vessel pixels as negative. Fig. 3 displays an example that demonstrates the

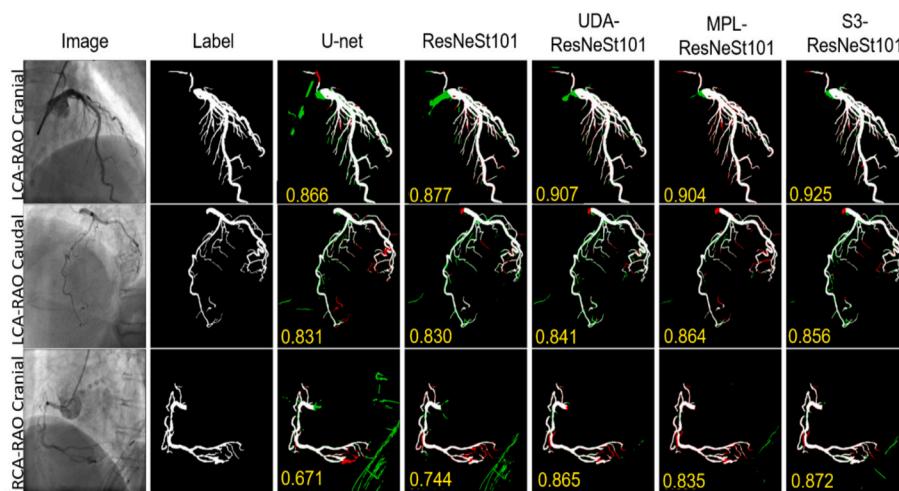


Fig. 3. Results for different types of supervised (U-Net and ResNeSt101) and semi-supervised models using ResNeSt101. (Color in each predicted result: Yellow Number - DSC; White pixel - true positive; Black pixel - true negative; Green pixel - false positive, vessel prediction in non-vessel label; Red pixel - false negative, non-vessel prediction in vessel label). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

performance metrics, calculated by comparing the predicted results against the labels. The white and black pixels in the model-predicted results represent the true positive and true negative cases, respectively, while the green and red pixels represent the false positive (vessel prediction in non-vessel label) and false negative (non-vessel prediction in vessel label) cases, respectively. Thus, the white and red pixels represent all pixels labeled as part of a vessel. Recall is the ratio of the number of white pixels to the number of white plus red pixels; it represents the proportion of correctly predicted vessel pixels in all pixels labeled as vessels, and is higher for a smaller fraction of red pixels. Conversely, the precision, which is the ratio of correctly predicted vessel pixels to all pixels predicted as vessels, has a higher value for a smaller fraction of green pixels. The threshold (TH) can be adjusted to favor either the recall or the precision. If we aim to improve the recall, the model will detect any linear object as a vessel, including a catheter or bone edge. Conversely, a higher precision requires a very strict vessel detection rule, thereby causing discontinuity of the vessel prediction. As a trade-off, we utilized the DSC, equal to the F1 score, which is the harmonic mean of the recall and precision, as the main outcome to confirm the model performance. Detailed discussion on the tradeoff between precision-recall is referred to Supplemental section and Supplemental Fig. S1.

3.2. Different semi-supervised methods using ResNeSt101

Besides mean teacher [14] and UDA [15], we compared different semi-supervised learning methods (Table 1) such as MPL [16], VAT [17], CPS [19], ST, and ST++ [20], all using ResNeSt101-UNet as the backbone. As shown in Table 1, compared with other existing semi-supervised methods, our proposed S3 differentiates others on the unique combination of consistency regularization and self-training,

enriched by the novel design of two-way pseudo-labeling, the strategic application of GDL/MSE loss, and the selection from the random augmentations in blur and contour (in Supplemental Table S1). Compared to the results of MT-UNet and UDA-UNet (in Table 2), using ResNeSt101-UNet as the backbone of MT and UDA (in Table 1) did not significantly improve the performance. That is, the advantage of the complex encoder is not entirely revealed in the consistent regularization-based method, demonstrating that solely using consistent regularization is insufficient for the interplay between teacher-student networks. In contrast, although MPL combined consistency regularization and pseudolabeling, the adoption of binary cross-entropy (BCE) as the loss function for the supervised learning degraded the performance than it did for the GDL of other methods. For a fair comparison, the vessel threshold for determining if a pixel belongs to a vessel was set as TH = 0.5. We observed that adopting GDL as a loss function would make the threshold value insignificant, and different threshold values will yield similar performances. However, the threshold was considered more sensitive when the BCE was used, as detailed in Supplemental section (a). Based on [31], we also ran paired-sample *t*-tests in DSC to analyze the statistical significance among three different cases, BCE/BCE and GDL/BCE, BCE/BCE and GDL/MSE, and GDL/BCE and GDL/MSE. By setting up the significance level as 0.05, the resulting p-values of all three cases are <0.001, demonstrating statistically significant improvement in adopting GDL/MSE than BCE/BCE and GDL/BCE. Thus, we utilize GDL/MSE as the loss function of the proposed S3 (in Supplemental Table S2).

Moreover, the principal design of MPL is to update the teacher network from the feedback of the student network, which does not provide further performance gain on the coronary artery segmentation. Thus, the S3 design does not consider the feedback between the student-teacher networks. Owing to the shape and boundary differences, the use of CPS or ST++ in this dataset was less effective than with other

Table 1

Comparison of different semi-supervised segmentation methods and their DSCs using ResNeSt101 as basic network (with 400 labels, 8952 unlabeled images).

Method	Loss function	Consistency regularization	Self-training	DSC
MT	GDL/BCE	Gaussian Noise	No	0.8788 ± 0.036
UDA	GDL/MSE	RandAugment	No	0.8895 ± 0.030
VAT [21]	BCE/KL Div.	Adversarial Perturbation	No	0.8826 ± 0.034
MPL	BCE/BCE	RandAugment	Pseudo-labeling with Feedback	0.8798 ± 0.033
MPLw/o feedback	BCE/BCE	RandAugment	Pseudo-labeling	0.8844 ± 0.033
CPS [29]	BCE/BCE	RandAugment	Cross Pseudo Supervision	0.8893 ± 0.034
ST	BCE/BCE	Strong Data Augmentations	Pseudo-labeling	0.8898 ± 0.033
ST++ [30]	BCE/BCE	Strong Data Augmentations	Select & Prioritize Reliable Image	0.8914 ± 0.033
S3	GDL/MSE	RandAugment	Two-way Pseudo-labeling	0.8999 ± 0.028

Table 2

Comparison in terms of 5 % labels case: 20 labeled and 8952 unlabeled images.

Method	Category	Pre-training	Recall	Precision	DSC
U-Net	Supervised	No	0.716 ± 0.121	0.780 ± 0.123	0.732 ± 0.081
ResNeSt101	Supervised	Yes	0.803 ± 0.075	0.827 ± 0.101	0.809 ± 0.065
ResNeSt101 non-pretrained	Supervised	No	0.661 ± 0.156	0.751 ± 0.109	0.688 ± 0.105
MT-UNet	Semi-supervised	No	0.820 ± 0.089	0.807 ± 0.101	0.807 ± 0.069
MT-ResNeSt101	Semi-supervised	Yes	0.838 ± 0.070	0.817 ± 0.103	0.822 ± 0.065
UDA-UNet	Semi-supervised	No	0.788 ± 0.110	0.840 ± 0.088	0.806 ± 0.078
UDA-ResNeSt101	Semi-supervised	Yes	0.818 ± 0.078	0.828 ± 0.074	0.818 ± 0.053
S3-UNet	Semi-supervised	No	0.746 ± 0.127	0.832 ± 0.099	0.776 ± 0.092
S3-ResNeSt101	Semi-supervised	Yes	0.828 ± 0.086	0.876 ± 0.071	0.847 ± 0.059
S3-ResNeSt101 non-pretrained	Semi-supervised	No	0.830 ± 0.064	0.822 ± 0.094	0.823 ± 0.063

semantic segmentation datasets, such as Cityscapes and Pascal VOC. The irregular shape and thickness of vessels in CAG images cause complex relationships between the pixel distributions, making it difficult to distinguish between the background and blood vessels. Thus, cross-consistency in CPS and model update iteration in ST++ have less reliable benefits than simply using two-way masking in S3 per pixel.

Fig. 3 shows examples of the results of the supervised and semi-supervised methods using ResNeSt101. The DSC of each sample from different models is indicated in the figures. In UNet LCA–RAO Cranial angiography, the prediction wrongly identified the catheter (green color, false positive) as vessels, unlike in the semi-supervised methods. In UNet/ResNeSt101 RCA–RAO Cranial angiography, the rib was wrongly identified as vessels (green color), which was rarely observed in the semi-supervised method. The supervised UNet and ResNeSt101 produced noticeable green pixels (false positive) that mis-predicted the non-vessel pixels as vessel, thus lowering the precision. In Fig. 4, we further compared the proposed S3-ResNeSt101 with UNet for a different number of labels. The S3 method showed a higher recall rate (fewer red pixels). Overall, based on preservation of continuity of the predicted vessels, the semi-supervised techniques improved the DSC.

3.3. Performance comparison under different percentages of labeled samples

When evaluating segmentation approaches that fall under the semi-supervised learning setting, standard practices assess their performance on more extreme annotated scenarios. Considering that most fully supervised contain fewer pixel-wise masks, e.g., in the public dataset [27]

such as STARE, DRIVE, and CHASE-DB1 containing 20 or 28 images, we considered 5 %, 10 %, and 100 % of labels case to train the proposed S3 model in Tables 2, 3, and 4 with MT [14] and UDA [15].

Since a deeper encoder (ResNeSt101) is considered in the U-Net's encoder, we compared pre-training using ImageNet or not (termed non-pre-trained) in both supervised and semi-supervised settings.

From the data-driven perspective, the semi-supervised S3 applied to a non-pre-trained U-Net-backbone network significantly improves from the supervised non-pre-trained U-Net-backbone network, e.g., DSC of non-pre-trained S3-ResNeSt101 0.823 enhanced from that of non-pre-trained supervised ResNeSt101 0.688 under 5 % case. With the information extracted from unlabeled data in the S3 scheme, the U-Net-backbone network, even without pre-training, still outperforms UNet, e.g., DSC of non-pre-trained S3-ResNeSt101 0.823 compared to that of S3-UNet 0.776 under 5 % case; that is, the proposed S3 adopts the unlabeled information well enough to bring out the advantage of deeper U-Net-backbone network. Furthermore, semi-supervised S3 using the pre-trained U-Net-backbone network has further performance gain, e.g., DSC of S3-ResNeSt101 improved to 0.847 under 5 % case. In addition, the same performance trend can be found in the sufficient (100 %) case.

Table 4 shows that compared to other semi-supervised methods, the proposed S3 method with U-Net improves the DSC from 0.846 of supervised U-Net to 0.888. Notably, the semi-supervised U-Net (considered a relatively simple model) learning from the data slightly outperformed the advanced supervised U-Net backbone networks (ResNeSt101) with pretraining. In addition, with the help of unlabeled data, S3-ResNeSt101 with only 5 % labels reaches a DSC of 0.847 in Table 2, and S3-UNet with only 10 % labels reaches a DSC of 0.846 in

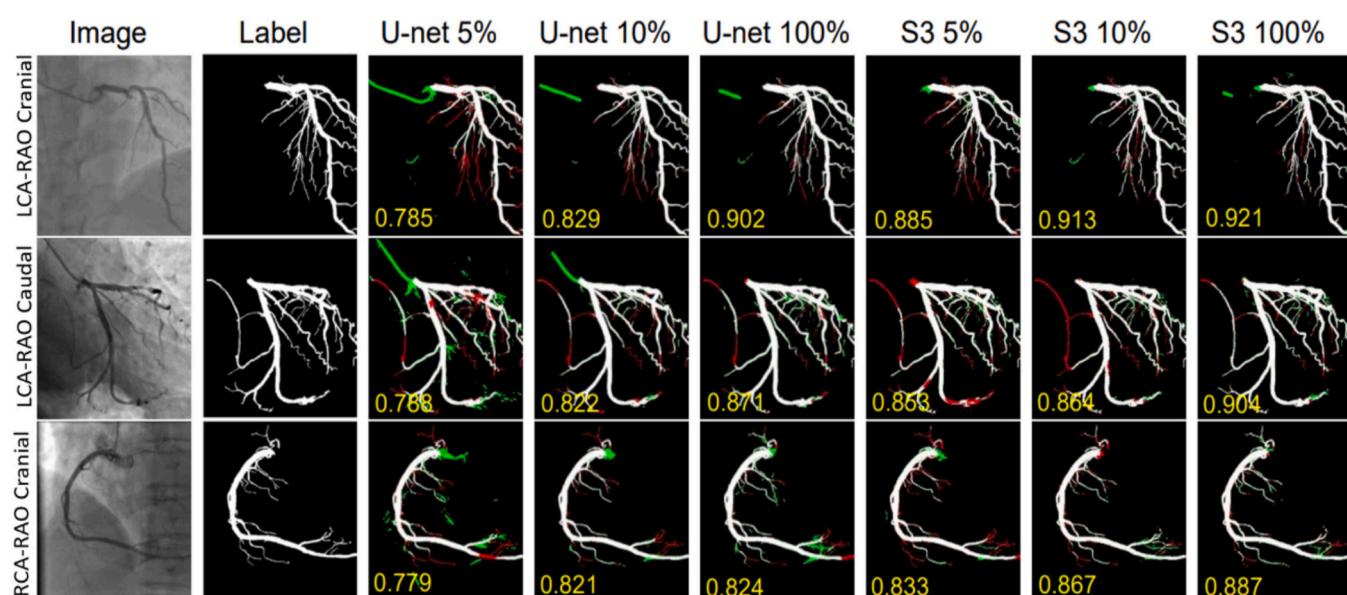


Fig. 4. Predicted result comparison between supervised U-Net and S3 using ResNeSt101 with different percentages of label number.

Table 3

Comparison in terms of 10 % labels case: 40 labeled and 8952 unlabeled images.

Method	Category	Pre-training	Recall	Precision	DSC
U-Net	Supervised	No	0.772 ± 0.103	0.808 ± 0.109	0.781 ± 0.076
ResNeSt101	Supervised	Yes	0.838 ± 0.062	0.845 ± 0.082	0.838 ± 0.052
ResNeSt101 non-pretrained	Supervised	No	0.737 ± 0.106	0.792 ± 0.116	0.754 ± 0.083
MT-UNet	Semi-supervised	No	0.852 ± 0.073	0.814 ± 0.010	0.827 ± 0.064
MT-ResNeSt101	Semi-supervised	Yes	0.820 ± 0.069	0.866 ± 0.083	0.838 ± 0.054
UDA-UNet	Semi-supervised	No	0.827 ± 0.094	0.848 ± 0.084	0.832 ± 0.068
UDA-ResNeSt101	Semi-supervised	Yes	0.839 ± 0.053	0.857 ± 0.071	0.845 ± 0.045
S3-UNet	Semi-supervised	No	0.823 ± 0.080	0.879 ± 0.073	0.846 ± 0.059
S3-ResNeSt101	Semi-supervised	Yes	0.839 ± 0.074	0.894 ± 0.069	0.862 ± 0.055
S3-ResNeSt101 non-pretrained	Semi-supervised	No	0.826 ± 0.079	0.879 ± 0.065	0.848 ± 0.055

Table 4

Comparison in terms of 100 % labels case: 400 labeled and 8952 unlabeled images.

Method	Category	Pre-training	Recall	Precision	DSC
U-Net	Supervised	No	0.847 ± 0.099	0.856 ± 0.071	0.846 ± 0.066
ResNeSt101	Supervised	Yes	0.898 ± 0.038	0.875 ± 0.054	0.885 ± 0.034
ResNeSt101 non-pretrained	Supervised	No	0.876 ± 0.053	0.877 ± 0.055	0.875 ± 0.039
MT-UNet	Semi-supervised	No	0.880 ± 0.063	0.882 ± 0.058	0.878 ± 0.043
MT-ResNeSt101	Semi-supervised	Yes	0.894 ± 0.041	0.885 ± 0.047	0.888 ± 0.031
UDA-UNet	Semi-supervised	No	0.882 ± 0.057	0.889 ± 0.053	0.883 ± 0.039
UDA-ResNeSt101	Semi-supervised	Yes	0.904 ± 0.039	0.877 ± 0.054	0.889 ± 0.035
S3-UNet	Semi-supervised	No	0.890 ± 0.052	0.891 ± 0.050	0.889 ± 0.035
S3-ResNeSt101	Semi-supervised	Yes	0.894 ± 0.041	0.908 ± 0.042	0.899 ± 0.028
S3-ResNeSt101 non-pretrained	Semi-supervised	No	0.884 ± 0.050	0.896 ± 0.060	0.889 ± 0.041

Table 3, which performed similarly to the supervised U-Net 0.846 with 100 % case in **Table 4**. For discussion on the choice of other U-Net backbone networks or TransUNet, we summarize the overall model selection trade-off in terms of model parameters, latency, and DSC in [Section 3.5](#).

3.4. Application to CAG film

CAG is performed using continuous and rapid angiography to observe the injection of contrast medium into the lumen of coronary arteries. In the resultant CAG film, the heartbeats and movements during the systolic and diastolic periods can cause background clutter and increase the difficulty of demarcating the vessels. We labeled 16 continuous frames from the LCA–RAO Cranial, LCA–RAO Caudal, and RCA–RAO Cranial angiographies and sampled every two frames for performance testing, as shown in [Fig. 5a](#), b, and c. The same set of labeled and unlabeled images used in the previous experiments was used for the supervised and semi-supervised training. We observed the segmentation performance of the angiography based on the training dataset, which was labeled from only one selected frame of each angle of angiography.

Compared with the supervised learning methods, the training dataset of the semi-supervised learning methods included additional unlabeled data, which provided extra information to distinguish the background noise from the vessels. [Fig. 5a](#) shows serious vascular prediction errors in the first two frames. The details of the full branches related to main arteries detected by S3-ResNeSt101 were better than that of the supervised UNet. For the LCA–RAO Caudal angiography, [Fig. 5b](#) reveals that the supervised UNet caused vessel discontinuity in frames 17–19. In addition, the proposed S3-ResNeSt101 method exhibited slight discontinuities; however, it predicted the appearance of vessels in the first few frames of [Fig. 5b](#) with better performance.

For the RCA–RAO Cranial angiography prediction shown in [Fig. 5c](#), UNet produced many fragmented and discontinuous vessel predictions, resulting in poor DSC. Further, [Fig. 5c](#) shows that the proposed S3-ResNeSt101 was more refined than supervised learning, especially for the main artery continuity. Overall, for the LCA–RAO Cranial angiography, the proposed S3-ResNeSt101 achieved an average DSC of 0.8645

± 0.012, which was better than that of supervised UNet (0.8420 ± 0.0286). S3-ResNeSt101 outperformed UNet in LCA–RAO Caudal angiography and RCA–RAO Cranial angiography as well (0.8496 ± 0.013, 0.8957 ± 0.008 vs. 0.8157 ± 0.0258, 0.7664 ± 0.0585, respectively).

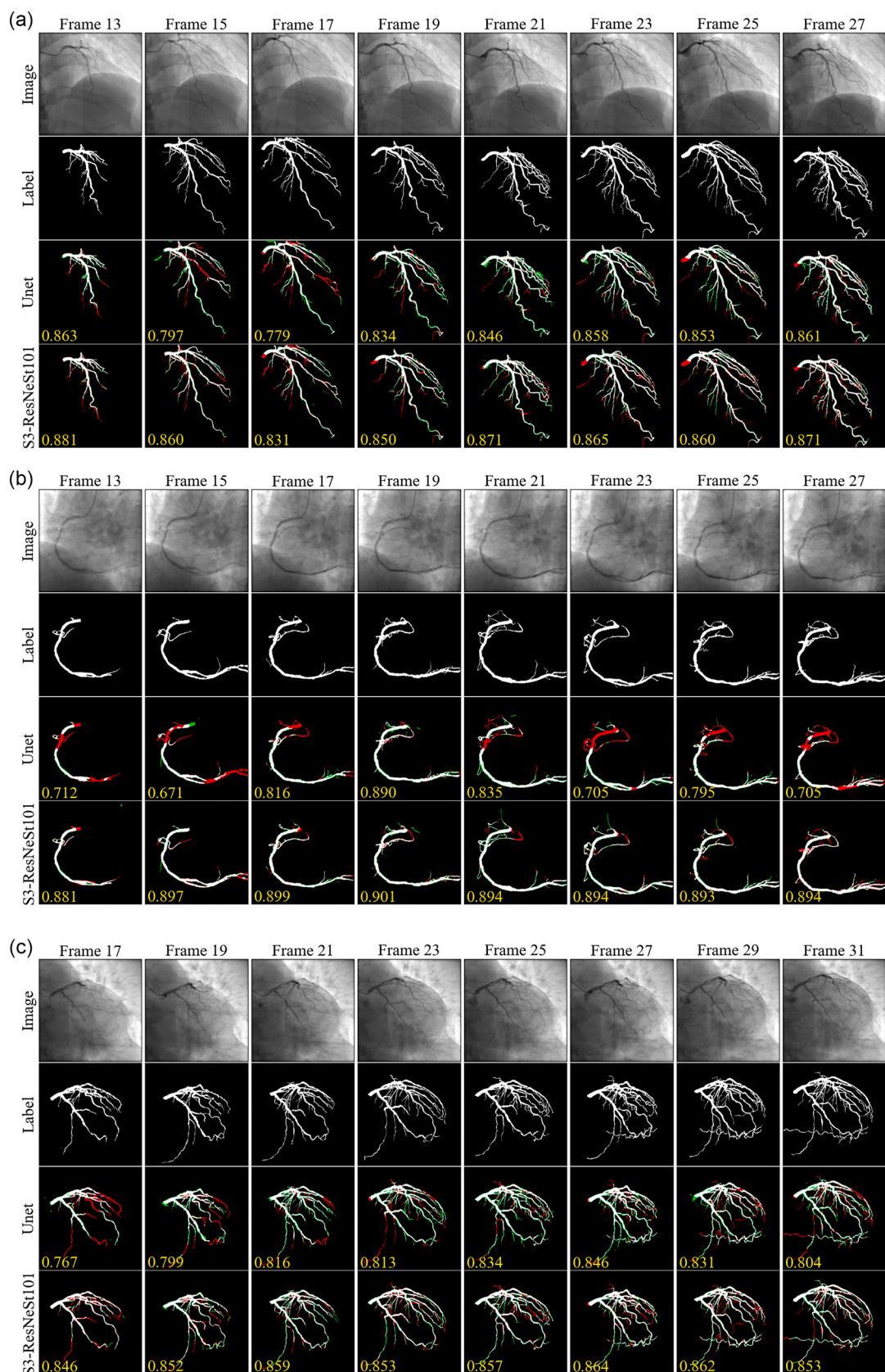
3.5. Model selection tradeoff

We compared the model parameters and latency for online prediction in **Table 5**. On the backbone network selection, notice that the deeper CNN-based network provided better performance for the supervised networks, and the same trend can be observed for the semi-supervised networks. Besides, we also compare with TransUNet as an advanced supervised learning method from the transformer-based network. As illustrated in **Table 5**, TransUNet performs slightly better than UNet but worse than CNN-based networks such as DenseNet, EfficientNet, and ResNeSt in the considered dataset under the same batch size setting. Notice that CNN models outperform some transformer-based models on small datasets because CNNs are more detail-oriented.

Owing to only the student network being utilized in the proposed S3 method for online prediction, the parameter number is similar to that of the supervised methods with the same basic network. Because the frame rate of the CAG film was 15 frames per second (fps), we utilized the model with an online prediction latency of fewer than 66.67 milliseconds (ms) (i.e., S3-ResNeSt101) as our baseline proposed scheme. This latency allows the display of the predicted result before the next frame input. We chose to compare S3-ResNeSt101 with S3-EfficientNetB5 instead of other versions of EfficientNet because these two models had a similar latency performance in our testing environment. In addition, the proposed S3 method had a better DSC with a slightly increasing online latency (10 ms) than the supervised method with the same basic network.

3.6. Further improvement of S3-ResNeSt101 and S3-EfficientNetB5

We tested different teacher–student model performances, specifically for the S3 method, as shown in **Table 6**. We set up the teacher network using either ResNeSt101 or EfficientB5, which have similar

**Fig. 5.** Application to the prediction of CAG films.

(a) Samples frames from LCA - RAO Cranial angiography.

(b) Samples frames from LCA - RAO Caudal angiography.

(c) Samples frames from RCA - RAO Cranial angiography.

Table 5
Model selection trade-off in terms of model parameters, latency, and DSC (with 400 labels, 8952 unlabeled images).

Method	Supervised				Semi-Supervised			
	UNet	DenseNet121-UNet	ResNeSt101-UNet	EfficientNetB5-UNet	TransUNet	S3-UNet	S3-ResNeSt101-UNet	S3-EfficientB5-UNet
DSC	0.8464 ± 0.066	0.8839 ± 0.031	0.8851 ± 0.034	0.8859 ± 0.031	0.86186 ± 0.369	0.8888 ± 0.035	0.8999 ± 0.028	0.8989 ± 0.028
Precision	0.8555 ± 0.071	0.8836 ± 0.048	0.8752 ± 0.054	0.8740 ± 0.052	0.87797 ± 0.052	0.8915 ± 0.050	0.9081 ± 0.042	0.9110 ± 0.044
Recall	0.8475 ± 0.099	0.8867 ± 0.041	0.8978 ± 0.038	0.9010 ± 0.040	0.85025 ± 0.057	0.8896 ± 0.052	0.8941 ± 0.041	0.8859 ± 0.050
Parameters (M)	13.39	13.60	55.26	31.22	105.91	13.39	55.26	31.22
Latency (ms)	8.02	36.83	56.31	56.72	60.36	9.26	66.73	67.56

Table 6
Further improvement on S3-ResNeSt101 and S3-EfficientNetB5.

Teacher model	Student model	DSC
ResNeSt101	ResNeSt101	0.8999 ± 0.0279
ResNeSt101	EfficientNetB6	0.9003 ± 0.0292
EfficientNetB5	EfficientNetB5	0.8989 ± 0.0276
EfficientNetB5	EfficientNetB6	0.8996 ± 0.0292

DSC-latency performances, and adopted EfficientB6 as a more complex student model, as suggested by the concept of noisy student method [26]. Replacing the student networks with EfficientNetB6 yielded the best DSC of 0.90034. Moreover, the performance of the teacher network using EfficientNetB5 was slightly worse than that using ResNeSt101.

4. Discussion

Accurate and precise CAG image segmentation is essential in the diagnosis of coronary artery disease. Semantic segmentation, which attempts to semantically identify the object that each pixel represents in the image, requires the model to make pixel-wise predictions. Given CAG films as a series of images with a higher number of input images, the expensive and human labor-intensive pixel-wise labeling that needs medical domain knowledge limits the possibility of constructing a sufficiently large dataset for supervised learning. In this study, a semi-supervised semantic segmentation S3 is investigated on coronary arteries to evaluate its benefits in terms of both performance and reducing the number of labels. With 9452 selected CAG frames, we trained the model by labeling only 5.3 % of the whole data and fully utilizing the rest. The results show that not only S3 with the help of unlabeled images outperforms a supervised network that has the same basic network using the same number of labeled images, but also S3-UNet with only 25 % of labeled images outperforms supervised U-Net with 100 % labeled images.

Our enhanced investigation into intra-surgery CAG continuous films highlights the dynamic capabilities of our S3 model when incorporating unlabeled data alongside labeled end-systolic/end-diastolic frames. This approach enables our model to excel in film segmentation prediction, adeptly handling the variability in vessel distribution, cardiac cycle phases, and contrast medium filling states, which may not directly align with those of the labeled frames. This adaptability is vividly illustrated in Fig. 3, where we observe that supervised learning models falter, unable to discern between essential elements (such as catheters and bone edges) and irrelevant features, with even minor discrepancies leading to inaccuracies. In contrast, our semi-supervised learning model benefits from a wealth of data, learning to differentiate real vascular structures from background noise, variations in the cardiac cycle, and motion artifacts effectively.

The proposed semi-supervised semantic segmentation method utilizes U-Net by replacing the backbone with ResNeSt and EfficientNet. Moreover, it utilizes the two-class generalized dice loss to balance the performance between recall and precision. The first teacher network was trained by designing a consistency mask for semantic segmentation. The consistency loss with confidence masking was calculated between two images with different data augmentations for pixels with a predicted probability exceeding the vessel and background thresholds, which helps to have a stable performance by discarding low-confidence predicted pixels in the loss calculation. By training the teacher network, the generated predictions were treated as pseudo-labels for the student network to further distinguish the features. The design of these semi-supervised methods increased the semantic segmentation smoothness. The prediction latency of the backbone with ResNeSt reached the same sampling frequency of 15 fps as the CAG film, and the performance of S3 demonstrated its ability to use smaller labeled datasets and generate better results.

In connection with our scheme with attention, the backbone network

of our proposed S3 adopted ResNeSt101, which is a split-attention network, performing channel-wise attention on different network branches to leverage their success in capturing cross-feature interactions and learning diverse representations. On the difference between weakly supervised learning and semi-supervised learning, in the learning challenge-wise, weakly supervised learning infers detailed information (e.g., pixel-level details for segmentation) from less detailed labels, while semi-supervised learning seeks to enhance learning from a limited labeled dataset by incorporating insights drawn from a larger unlabeled dataset. On the other hand, applicability-wise, weakly supervised learning is often chosen when the available labels are coarse or imprecise but still informative. In contrast, semi-supervised learning is preferred when some high-quality labeled data are available, but more data is needed to improve learning outcomes without the additional cost of labeling. Thus, in our problem, we have high-quality but limited labeled data and a larger unlabeled dataset, which is more suitable for the scenario for semi-supervised learning.

Moreover, the semi-supervised model utilizes the interconnectedness of frames within CAG films to extract crucial information for artifact elimination. This not only showcases the superiority of semi-supervised learning over traditional supervised methods in dealing with the complexities of real-world medical imaging but also underlines its potential in refining our understanding of what constitutes genuine vascular imagery as opposed to mere noise, even amidst the challenges posed by cardiac movement.

In clinical settings, coronary angiography is usually obtained at a speed of 15 frames per second, aligning with heart rates around 80 beats per minute. This necessitates a latency of approximately 67 ms to merit the need for real-time processing, ensuring that our approach is perfectly aligned with the operational demands of cardiac procedures. Our proposed S3 model, as demonstrated in [Table 5](#) and Fig. S4, achieves this critical latency threshold while maintaining the highest DSC performance for accuracy. This balance of speed and precision underscores our method's potential in enhancing PCI's safety and effectiveness.

Building upon the existing coronary angiography framework, DL-assisted segmentation offers an advanced approach by generating a digital, motion-compensated overlay. This overlay, when superimposed on live fluoroscopic images, facilitates the visualization of the coronary tree, thereby enabling the precise navigation of wires and devices during the procedure without necessitating additional contrast medium or radiation exposure. This methodology inherently diminishes the need for repeated angiographies, reducing both contrast agent and radiation doses.

Drawing parallels to traditional image processing methods, such as the Dynamic Coronary Roadmap system—which has been shown to decrease the use of contrast medium and radiation by providing clearer visualization of coronary artery paths [28]—our DL-assisted approach extends these benefits. By integrating deep learning elements, it is adept at handling a wider range of medical scenarios and variations, further optimizing the procedural efficiency and safety. Our method not only aligns with the goals of reducing radiation and contrast doses but also enhances adaptability and precision in complex clinical settings.

This is our first few attempts toward reaching AI-assisted medicine. We only evaluate the model performance from statistical and technical aspects. In future work, we could apply our segmentation model to other medical parameters, such as stenosis and flow status. Through further development and clinical trials, we can keep our operating costs at a reasonable level while providing better quality healthcare.

5. Conclusion

CAG films are a series of images, therefore, the expensive and labor-intensive pixel-wise labeling requiring medical domain knowledge limits the possibility of constructing a sufficiently large dataset for supervised learning. In this study, semi-supervised semantic segmentation was applied to coronary arteries to evaluate its benefits in terms of both

performance and reduction in the label count. With 9452 selected CAG frames, we trained the model by labeling only 5.3 % of the entire data and fully utilizing the rest unlabeled images. The results showed that S3, using the same number of labeled images and some unlabeled images, outperformed a supervised network with the same basic network. Further, S3-ResNeSt101 with only 5 % of labeled images achieved similar performance as supervised UNet with 100 % labeled images. The prediction latency of S3-ResNeSt101 reached the frame rate of the CAG film (i.e., 15 fps), and the S3 performance demonstrated its ability to use smaller labeled datasets and generate better results. Additional experiments on clinical CAG films demonstrated that by adopting unlabeled data from the remaining unlabeled frames other than the labeled end-systolic/-diastolic frames, the static S3 model can achieve superior film segmentation prediction, even if the vessel distribution, phase of cardiac cycle, and contrast medium filling status in the film do not coincide with those of the labeled images. We concluded that all these works may provide a dynamic roadmap for PCI guidance, and the radiation and contrast doses might be further reduced.

Sources of funding

This work was supported in part by the National Science and Technology Council (NSTC) under Grant 109-2314-B-002-257-MY3, Grant 111-2221-E-027-055-MY2 and Grant 112-2314-B-002-192-MY3.

CRediT authorship contribution statement

Chih-Kuo Lee: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Data curation. **Jhen-Wei Hong:** Investigation, Formal analysis. **Chia-Ling Wu:** Software, Project administration, Investigation, Formal analysis. **Jia-Ming Hou:** Software, Project administration, Methodology, Investigation, Formal analysis. **Yen-An Lin:** Validation, Software, Investigation, Formal analysis. **Kuan-Chih Huang:** Investigation, Data curation, Conceptualization. **Po-Hsuan Tseng:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors have no conflict of interest to disclose.

Acknowledgments

The authors sincerely thank the team members of the medical imaging CAG dataset project, which is funded by the National Science and Technology Council. Their support has been instrumental in enabling this research. Our team members Zheng-Zhang Li, Zhi-Gao Qiu, Yeh-Chun Lu, and Hsiao-Shan Chen also contribute to the CAG data construction and other relevant research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.artmed.2024.102888>.

References

- [1] Mahendiran T, Thanou D, Senouf O, Meier D, Dayer N, Aminfar F, et al. Deep learning-based prediction of future myocardial infarction using invasive coronary angiography: a feasibility study. *Open Heart* 2023;10:e002237.
- [2] Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998. p. 130–7.

- [3] Hao D, Ding S, Qiu L, Lv Y, Fei B, Zhu Y, et al. Sequential vessel segmentation via deep channel attention network. *Neural Netw* 2020;128:172–87.
- [4] Mou L, Zhao Y, Fu H, Liu Y, Cheng J, Zheng Y, et al. CS2-Net: deep learning segmentation of curvilinear structures in medical imaging. *Med Image Anal* 2021; 67:101874.
- [5] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: transformers make strong encoders for medical image segmentation. In: arXiv preprint arXiv: 210204306; 2021.
- [6] Li L, Liu Q, Shi X, Wei Y, Li H, Xiao H. UCFilTransNet: cross-filtering transformer-based network for CT image segmentation. *Expert Syst Appl* 2024;238:121717.
- [7] Qin B, Mao H, Liu Y, Zhao J, Lv Y, Zhu Y, et al. Robust pca unrolling network for super-resolution vessel extraction in x-ray coronary angiography. *IEEE Trans Med Imaging* 2022;41:3087–98.
- [8] Yang S, Kweon J, Kim Y-H. Major vessel segmentation on X-ray coronary angiography using deep networks with a novel penalty loss function. 2019.
- [9] Yang S, Kweon J, Roh J-H, Lee J-H, Kang H, Park L-J, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep* 2019;9: 16897. <https://doi.org/10.1038/s41598-019-53254-7>.
- [10] Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, et al. Multichannel fully convolutional network for coronary artery segmentation in X-ray angiograms. *IEEE Access* 2018;6:44635–43.
- [11] Kassim Y, Glinskii O, Glinsky V, Huxley V, Guidoboni G, Palaniappan K. Deep U-net regression and hand-crafted feature fusion for accurate blood vessel segmentation. In: International conference on image processing; 2019. p. 1445–9.
- [12] Cervantes-Sanchez F, Cruz-Aceves I, Hernandez-Aguirre A, Hernandez-Gonzalez MA, Solorio-Meza SE. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Appl Sci* 2019;9:5507.
- [13] Carballal A, Nóvoa F, Fernandez-Lozano C, García-Guimaraes M, Aldama G, Vázquez Rodríguez J, et al. Automatic multiscale vascular image segmentation algorithm for coronary angiography. *Biomed Signal Process Control* 2018;46. <https://doi.org/10.1016/j.bspc.2018.06.007>.
- [14] Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in neural information processing systems. Curran Associates, Inc.; 2017.
- [15] Xie Q, Dai Z, Hovy E, Luong T, Le Q. Unsupervised data augmentation for consistency training. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in neural information processing systems. Curran Associates, Inc.; 2020. p. 6256–68.
- [16] Pham H, Dai Z, Xie Q, Le QV. Meta pseudo labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021. p. 11557–68.
- [17] Miyato T, Maeda S-i, Koyama M, Ishii S. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Trans Pattern Anal Mach Intell* 2018;41:1979–93.
- [18] Zoph B, Ghiasi G, Lin T-Y, Cui Y, Liu H, Cubuk ED, et al. Rethinking pre-training and self-training. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in neural information processing systems. Curran Associates, Inc.; 2020. p. 3833–45.
- [19] Chen X, Yuan Y, Zeng G, Wang J. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 2613–22.
- [20] Yang L, Zhuo W, Qi L, Shi Y, Gao Y. St++: make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p. 4268–77.
- [21] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Kamalika C, Ruslan S, editors. International conference on machine learning: PMLR; 2019. p. 6105–14.
- [22] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016. p. 770–8.
- [23] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
- [24] Zhang H, Wu C, Zhang Z, Zhu Y, Zhang Z, Lin H, et al. ResNeSt: a split-attention networks. CoRR; 2020 (<https://arxiv.org/abs/2004.08955>).
- [25] Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Proceedings 3. Cham: Springer International Publishing; 2017. p. 240–8.
- [26] Xie Q, Luong M-T, Hovy E, Le QV. Self-training with noisy student improves ImageNet classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 10687–98.
- [27] Liu W, Yang H, Tian T, Cao Z, Pan X, Xu W, et al. Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. *IEEE J Biomed Health Inform* 2022;26:4623–34. <https://doi.org/10.1109/jbhi.2022.3188710>.
- [28] Piayda K, Phinicardies R, Afzal S, Veulemans V, Jung C, Bönnér F, et al. Dynamic coronary roadmap in percutaneous coronary intervention: results from an open-label, randomized trial. *JACC Cardiovasc Interv* 2021;14:2523–5.
- [29] Xie Q, Luong M-T, Hovy E, Le QV. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2021. p. 2613–22.
- [30] Yang L, Zhuo W, Qi L, Shi Y, Gao Y. St++: make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2022. p. 4268–77.
- [31] Shen N, Xu T, Bian Z, Huang S, Mu F, Huang B, Xiao Y, Li J. SCANet: a unified semi-supervised learning framework for vessel segmentation. *IEEE Trans Med Imaging* 2023;42:2476–89.