StenosisSeg: Automatic Stenosis Segmentation for Coronary Artery Disease

Tim Jing
Stanford University
timjing@stanford.edu

Andy Wang
Stanford University
wangandy@stanford.edu

Emily Liu
Stanford University
emily712@stanford.edu

Abstract

Identifying stenotic blood vessel narrowing caused by underlying coronary artery disease (CAD) has been of significant interest due to its profound clinical impact. We present a flexible, generalizable stenosis segmentation model for X-ray Coronary Angiography (XCA) images with potential for real-time clinical diagnostics. We build upon Mask R-CNN and incorporate a pseudo-labeling data augmentation procedure. Specifically, a supervised base Mask R-CNN model trained on labeled stenosis images is subsequently used to propose labels on other unlabeled XCA images, which can be then fed back into a final pseudo-label model for training. Our pseudo-label augmentation resulted in a modest boost to the F1 evaluation metric on the ARCADE dataset, an expert-labeled XCA stenosis image set. More importantly, the pseudo-labeling procedure enables the incorporation of XCA images in other datasets coming from diverse hospitals, equipment, and patient populations during training. All of our models significantly outperformed the baseline F1 scores achieved by YOLOv8n-seg, with our best model achieving a test F1 score of 0.51. Current SOTA architecture on the ARCADE dataset achieves an F1 score of 0.54 with significantly more compute. Our code can be found at: https://github.com/liuemi001/ArterySeg.

1. Introduction

Coronary artery disease (CAD) is a cardiovascular condition caused by a buildup of atherosclerotic plaque in the coronary arteries, and is one of the leading causes of death worldwide, affecting about 5% of the global population [20]. The plaque buildup causes a narrowing in these vessels, referred to as stenosis, which can endanger the patient by reducing blood supply to the heart.

One of the most commonly used tools to diagnose coronary artery disease is X-ray Coronary Angiography (XCA), wherein a contrast agent is injected into a patient's coronary arteries via a catheter, and subsequent X-ray imaging is performed to visualize the arteries. Although some argue that Computed Tomography Coronary Angiography (CTCA) is a more valuable tool in assessment of coronary artery disease due to its ability to capture 3D structures [6], we chose to work with XCA due to its superior diagnostic ability in severe and obstructive CAD cases [18].

Traditional vessel segmentation and stenotic lesion segmentation techniques are labor-intensive and time-consuming, making them unscalable to large amounts of data. **Several deep learning methods have been developed for general coronary artery segmentation, yet, to our knowledge, no such established methods are commonly used in a clinical setting to automate stenosis segmentation in XCA images using deep learning**. Thus, constructing a deep learning computer vision system to augment stenosis segmentation with high precision and computational efficiency for clinical deployment would be extremely valuable. Such an automatic stenosis detection system could serve as an assistant to radiologists, performing a "first pass" over the huge number of XCA images produced per year in order to reduce the workload of radiologists. This would help streamline clinical pipelines, and also potentially ameliorate the high rates of burnout experienced by radiologists, largely due to substantially increased workloads in the past 20 years. [2]

To this end, we develop an algorithm which takes an XCA image of any size as input, and uses a Mask R-CNN model to produce a predicted binary segmentation mask indicating regions of stenosis.

2. Related Works

Previous works have demonstrated the potential of deep learning for the task of coronary artery and stenosis localization using CTCA data. For example, Li et al. were able to achieve an F1 score of 0.775 on the task of stenosis segmentation [13], while Huang et al. report a Dice score of 0.71 for artery segmentation [9], both with U-Net-based architectures. However, very few works have utilized XCA

images, and most of these works have found success only on artery segmentation [5] [21].

There are very few released methods on stenosis detection, one of which was developed by Du et al. and is called DeepDiscern [4]. DeepDiscern uses two parallel deep neural networks—one that detects artery segments, and one that detects lesions–to generate high-level diagnostic information. The model achieved an F1 score of 0.829 on the stenotic lesion detection task. However, this work produces bounding boxes as their output, which are much less precise than segmentation masks, and can be more ambiguous. We purport that pixelwise instance segmentation of stenotic arteries would provide greater clinical value due to its greater precision, which is why we choose pixel-wise stenosis segmentation as our objective. The different objective of DeepDiscern (object detection vs instance segmentation), coupled with the fact that neither their model or dataset are publicly available, means that we are unable to directly compare our results with their evaluation metrics.

Currently, the most accessible dataset and benchmark for coronary artery stenosis segmentation from XCA images is the publicly released Automatic Region-based Coronary Artery Disease Diagnostics using X-ray angiography images (ARCADE) dataset, which contains 1500 XCA images with segmentation annotations of stenotic regions [18]. The annotations are provided in both YOLO and COCO formats, two established formats for representing segmentation masks. When releasing the dataset, the authors also trained a YOLOv8 model on the stenosis dataset as a public baseline and were able to achieve an F1 score of 0.38.

A number of other models have made use of the ARCADE dataset. Some, like Pokhrel et al. [17], have made use of the portion of the ARCADE dataset that contains labels for different segments of the coronary arteries, rather than for stenoses. Others have tackled the same task as us, the stenosis segmentation task, including StenUNet [14], Bilal et al. [1], and SSASS [11], which is the current SOTA, achieving a test F1 score of 0.54. However, many of these models use a YOLO-based architecture. Our work explores the usage of Mask R-CNN for stenosis segmentation, which has not been done before to our knowledge.

Another publicly released dataset for stenosis detection in XCA images is one published by Danilov et al. [3], containing 8325 images from 100 patients who underwent XCA imaging at the Research Institute for Complex Problems of Cardiovascular Diseases (Kemerovo, Russia). Although this dataset contains only bounding box annotations of stenotic regions, making the results in [3] not directly comparable to ours, it is still useful as an additional source of XCA images containing stenoses. We opt to incorporate this dataset in our training set, as described in Section 3.4, because we felt it could be helpful for the generalizability of our models.

3. Methods and Technical Approach

3.1. Baselines

As a baseline, we finetune YOLOv8 nano, a well established object detection and image segmentation model, on the raw ARCADE stenosis dataset. Specifically, we fine-tuned YOLOv8n-seg, which was pretrained on COCO 2017 and whose architecture is identical to YOLOv8 except that the object detection head of YOLOv8 is replaced with a segmentation head.

YOLO is a single-stage object detector, and instead of using region proposals to locate objects like in R-CNNs, YOLO runs a single convolutional network over the complete image to directly predict bounding boxes and class probabilities in a single, unified detection pipeline [19]. Specifically, YOLO divides the image into an $S \times S$ grid. Then, each of those grid cells predicts several bounding boxes and associated confidence scores, plus conditional class probabilities for that cell. The confidence score represents both how likely that there is an object in the box, as well as how accurate the bounds of the box are, and is defined as:

$$\text{Conf} = Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \tag{1}$$

During inference, the confidence scores for each box are multiplied by the class probabilities, yielding class confidence scores for each box (Figure 1). One important strength of the YOLO model is that since it takes in and considers the whole image, it can reason globally about the image and receive context from outside of the region or cell of interest.

For instance segmentation using YOLOv8n-seg, the YOLO architecture is modified slightly to include a small fully connected network called Proto, which generates segmentation masks [10].

3.2. Mask R-CNN

In addition to our baseline YOLO model, we propose three Mask R-CNN-based methods.

We first propose simply finetuning pre-trained Mask R-CNN ResNet-50 FPN [8] on the ARCADE stenosis dataset. We call this model **StenosisSeg-base**. Mask R-CNN is a two-stage object detector utilizing an initial ResNet backbone with a subsequent Region Proposal Network and

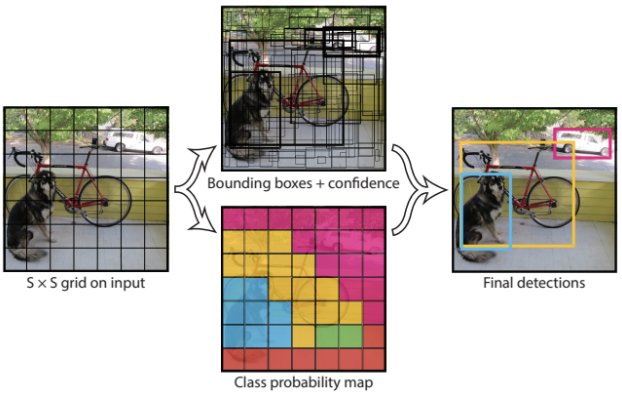

Figure 1. Diagram from Redmon et al. [19] of YOLO's unified detection network, demonstrating the synthesis of bounding box confidence scores and class probabilities for each of the $S \times S$ grid cells.

RoI pooling layer to identify the regions of the feature map to perform predictions on. When making predictions and computing losses, Mask R-CNN extends Faster R-CNN by incorporating a small network branch that predicts a segmentation mask for each RoI in parallel to the bounding box prediction that Faster R-CNN originally computes [8]. Specifically, we believe Mask R-CNN outperforms YOLO architecture because of its aforementioned ability to generate region proposals, which is notably absent in YOLO, hence the name [19]. While dividing images into cells and rapidly producing classifications lends well to real-time processing, when more compute and time can be afforded in the case of stenosis segmentation, we believe more complex region proposals have the edge.

In addition, Mask R-CNN implements RoIAlign, which helps decouple the mask prediction process from classification. Furthermore, in a task like image segmentation, the precise mapping of masks on small features maps to the original image afforded by RoIAlign is critical.

Specifically, RoIAlign does not quantize the floating-point RoI into the discrete feature map, which can lead to error. Instead, RoI Align retains the floating-point accuracy and then performs bilinear interpolation to "fill" the discrete boxes with a max or average value [8]. From Figure 2,

$$f(x,y) = \sum_{i,j=1}^{2} f_{i,j} \max(0, 1-|x-x_i|) \max(0, 1-|y-y_i|) \tag{2}$$

Thus, we hypothesize Mask R-CNN performs better in our stenosis segmentation task.

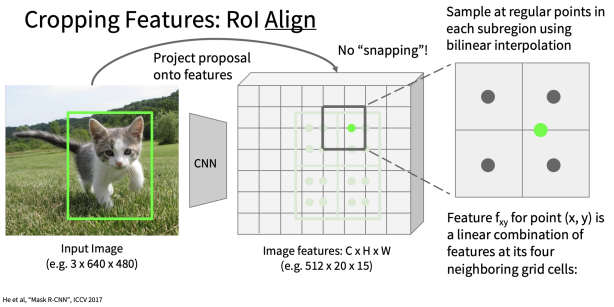

Figure 2. Illustration from Stanford University's CS 231N course [12] of RoIAlign's methodology.

The input image size of the Mask R-CNN architecture is 224x224, so we resize our images from 512x512 to 224x224 before model training. Finding that this simple training procedure results in significant overfitting over 50 epochs, while still not surpassing the ARCADE SOTA on validation set, we look into further data augmentation procedures.

3.3. Data Augmentation

The ARCADE dataset [18] consists of two different sets of 1000 CAD training images, one of which is labeled with stenosis segmentation masks (the "stenosis" dataset), and the other of which is labeled only with segmentation masks of different branches of the coronary arteries (the "syntax" dataset).

Our StenosisSeg-base model made use of only the "stenosis" dataset, but as the SSASS team [11] observed, while the syntax dataset is not explicitly labeled for stenoses, there are still stenoses present in each example of the syntax dataset. Thus, we make use of the syntax dataset as well by generating "pseudo-labels" of stenotic regions by running inference with StenosisSeg-base, following a similar procedure as the authors of the SSASS model [11]. The syntax mask and image combinations from inference are then concatenated with the initial stenosis dataset. We then train another Mask R-CNN model jointly on the combined stenosis and pseudo-labeled syntax dataset and call this model **StenosisSeg-pl** (see Figure 3).

3.4. Exploring Other Datasets

Additionally, to further augment our dataset and improve our models' generalizability outside the ARCADE dataset, we also explore incorporating another angiography dataset for stenosis detection introduced by Danilov et al. [3]. We choose not to train on the annotations in this dataset

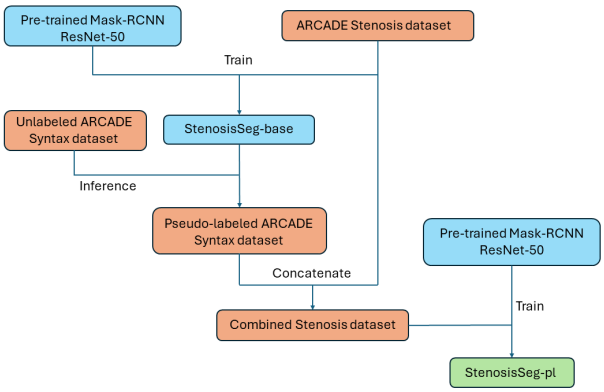

Figure 3. Schematic illustrating our data augmentation procedure, inspired by [11].

directly, as this dataset contains only bounding box annotations for stenotic regions and not segmentation masks, and converting bounding box labels to masks naively by labeling all pixels within the box as stenotic would introduce a significant amount of noise (many pixels not even within a vessel would be falsely labeled as stenotic). Instead, we opt to generate our own labels using the same pseudo-labeling procedure as described in Section 3.3.

4. Dataset and Features

The main dataset used, the ARCADE dataset, is an XCA image set labeled by medical experts spanning 1500 patients with a median age of 60. The images were obtained from a study cohort at the Research Institute of Cardiology and Internal Diseases (Almaty, Kazakhstan). Note that while each patient generated 60-120 frames, 0-12 frames were selected based on optimal contrast filling within the arteries and minimal blurriness among other properties to maximize training efficiency. In a direct clinical pipeline, this preprocessing would be necessary to ascertain live predictions from an XCA imaging machine [18].

The ARCADE dataset consists of two separate parts: the stenosis dataset and the syntax dataset. The stenosis dataset is the one that we primarily investigated and trained our models on. It consists of 1000 training images, 200 validation images, and 300 test images, all with associated bounding box and segmentation mask annotations in YOLO format. The second dataset, the syntax dataset, consists of images labeled with segmentation masks of different branches/regions of the coronary arteries. This dataset is also procured from patients with coronary artery disease. The syntax dataset is the same size as the stenosis dataset, with 1000 training images, 200 validation images, and 300 test images. All the images in these two datasets have dimensions of 512x512.

In order to prepare the dataset for Mask R-CNN training, we preprocessed the YOLO annotations, which consisted of polygon coordinates, into binary segmentation masks.

The second dataset incorporated from Danilov et al. [3] included 8325 XCA images. While bounding box annotations of stenotic regions were available, we used the raw images only for our pseudo-labeling pipeline.

5. Results and Discussion

5.1. Experimental Setup

We used batch size of 4 and an SGD optimizer with learning rate = 0.005, momentum = 0.9, and weight decay of 5×10^{-4}, and performed all training with a single NVIDIA T4 GPU. We chose these hyperparameters because we found that all our models were able to reach a very low training loss and very high training F1 score within 50 epochs or several hours using these hyperparameters. We used the publicly available pre-trained Mask R-CNN ResNet-50-FPN checkpoint for all Mask R-CNN fine-tuning experiments.

In order to evaluate performance of our models, we first threshold the segmentation masks produced by our Mask-RCNN-based models, which are not binary, but instead have soft values ranging from 0 to 1, representing the model's confidence that a particular pixel contains a stenotic vessel. Pixels given a score above a confidence threshold (which we tune using our validation set) are set to 1, while all other pixels are set to 0, in order to construct a binary mask of predicted stenotic regions. We then combine all predicted stenosis masks by performing a logical OR operation in order to produce one aggregated predicted mask. We use the same procedure to obtain an aggregated ground truth mask. We then compare the aggregated ground truth and predicted masks pixel-wise to calculate pixelwise F1 scores (a.k.a Dice similarity coefficients) [15].

$$F_1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{TP}}{\text{TP} + 0.5(\text{FP} + \text{FN})} \quad (3)$$

We choose our confidence threshold for each model by evaluating each model with all possible confidence thresholds (0-1) in increments of 0.05, and choosing the confidence threshold which yields the highest validation F1 score. Our proposed methods will be benchmarked against the baseline YOLOv8 results reported by the authors of ARCADE as well as our own baseline as described above.

Model	Val F1	Test F1
YOLOv8 (from [18])	-	0.38
YOLOv8n-seg	0.35	0.31
StenosisSeg-base	0.54	**0.51**
StenosisSeg-pl	**0.55**	**0.51**
StenosisSeg-other	0.49	0.47

Table 1. Validation and test set F1 scores of our methods. We also report metrics for two baseline methods: a YOLOv8-seg model trained by [18], and a YOLOv8n-seg we trained ourselves.

5.2. Baselines - YOLOv8

As a baseline, we fine-tuned YOLOv8n-seg on the ARCADE stenosis training set (n=1000) for 50 epochs and achieved an F1 score of 0.35 on the validation set and an F1 score of 0.31 on the test set (Table 5.1). The authors of ARCADE reported that their baseline YOLOv8 model achieved an F1 score of 0.38 [18]. The ARCADE team's slightly higher performance is likely due to their usage of a larger YOLO checkpoint (versus the "nano" checkpoint that we used), as well as a more thorough hyperparameter tuning process.

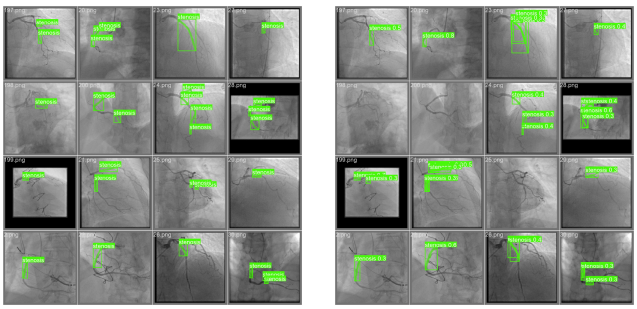

Figure 4. Left: Labels for a batch of validation images. Right: Predictions from baseline YOLOv8-seg for the same validation images at conf=0.01.

While some images are labeled accurately, the model was unable to segment certain images and provided more predictions than expected in others (Figure 4).

5.3. StenosisSeg-base Results

We trained a Mask R-CNN on the training set for 50 epochs. We found that although there was considerable overfitting by the 50th epoch (training F1 score was considerably higher, at around 0.80), the 50-epoch model still performed better on the validation set than all previous checkpoints. This checkpoint achieved the best validation F1 score of 0.54 at a confidence threshold of 0.55. When evaluated on the test set with the same confidence threshold, the model achieved an F1 score of 0.51 (Table 5.1).

Qualitatively, we also find that StenosisSeg-base outputs reasonably high-quality stenosis masks (Figure 5), validating our decision to use this model to generate pseudo-labels for further training. However, we do note that StenosisSeg-base is far from perfect, evident from Figure 6 and Figure 7, where StenosisSeg-base incorrectly identifies healthy vessels as stenotic or fails to identify the correct region of stenosis altogether, respectively. Additionally, it is interesting to note that sometimes when StenosisSeg-base correctly identifies the stenotic region in the image, it identifies multiple instances of stenosis when there is only one. 5

5.4. Data Augmentation Results - StenosisSeg-pl

As described in Section 3.3, we trained another Mask R-CNN on the combined dataset, which consisted of the original stenosis dataset as well as the syntax dataset pseudo-labeled with StenosisSeg-base (see Figure 3). We found that this model began to overfit at a much earlier epoch than the base Mask R-CNN model, likely due to the fact that the size of the dataset was double that of StenosisSeg-base. After evaluating several different checkpoints at different epochs on the ARCADE stenosis validation set, the epoch 30 checkpoint was found to have the best performance. We then did the same sweep of confidence threshold values on this checkpoint to find that this model achieved the best validation F1 score of 0.55 at a confidence threshold of 0.5. When evaluated on the test set, the model achieved an F1 score of 0.51 (Table 5.1).

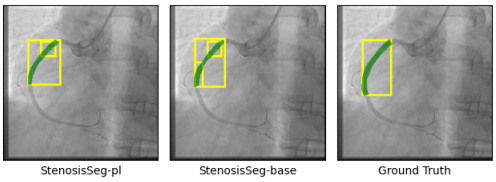

Figure 5. StenosisSeg-base predicted stenosis masks (middle) are qualitatively reasonably similar to ground truth stenosis labels (right), validating the use of StenosisSeg-base predictions as pseudo-labels. Stenosis-pl (left) also performs very well compared to ground truth. Each yellow bounding box in the ground truth labels represents a different predicted stenosis.

5.5. Training with Additional Datasets - StenosisSeg-other

In the spirit of improving generalizability, we experiment with incorporating data from a more diverse set of sources. We curate a combined dataset consisting of 1000 labeled examples from the ARCADE stenosis dataset,

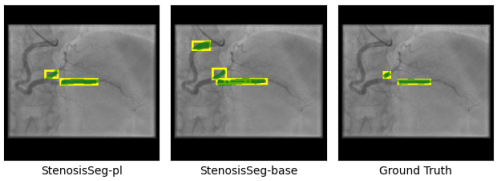

Figure 6. StenosisSeg-base predicted stenosis masks (middle) may predict additional stenotic regions compared to the ground truth (right). Stenosis-pl (left) correctly identifies the stenotic regions.

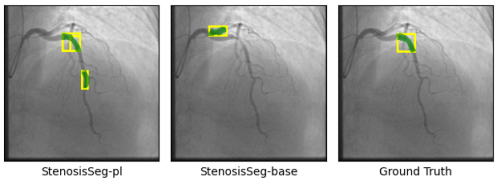

Figure 7. StenosisSeg-base predicted stenosis masks (middle) might fail to capture the correct region of stenosis compared to the ground truth (right). In addition, Stenosis-pl (left) has a tendency to predict additional stenotic regions.

1000 pseudo-labeled examples from the ARCADE syntax dataset, and 1000 pseudo-labeled examples from Danilov et al. [3]. The 1000 examples from Danilov et al. [3] were randomly subsampled from the original dataset of 8325 images in order to maintain balance between our three data sources, as we hypothesized that a training dataset still consisting mostly of images from ARCADE would perform better on the ARCADE validation and test sets.

Following the same procedure as our other Mask R-CNN based models, we fine-tune a pre-trained Mask R-CNN for a total of 30 epochs and evaluate checkpoints every 5 epochs on the validation set. We call the resulting model **StenosisSeg-other**. We find that this model achieved maximum performance after 10 epochs of training and at a confidence threshold of 0.65, at which it achieves an F1 score of 0.493. Running inference of the 10 epoch checkpoint with a confidence threshold of 0.65 on the test set yields an F1 score of 0.47 (Table 5.1).

While the F1 score on the ARCADE test set decreased, we hypothesize this was due to introducing images outside of the dataset. However, because not all images in the clinical setting will resemble ARCADE, StenosisSeg-other is still a meaningful step toward developing a general stenosis segmentation model.

6. Conclusion and Future Work

6.1. Conclusion

We introduce StenosisSeg, a new model built upon Mask R-CNN designed to segment stenotic lesions in patients with coronary artery disease (CAD). Specifically, this model operates on X-ray Coronoary Angiography (XCA) instead of Computed Tomography Coronary Angiography (CTCA), as XCA retains superior diagnostic ability in severe cases [18].

The pseudo-labeling data augmentation pipeline (StenosisSeg-pl) demonstrated potential compared to the vanilla Mask R-CNN model (StenosisSeg-base). While test F1 scores were very similar when rounded to two decimals (0.51 vs. 0.51), StenosisSeg-pl did achieve a higher F1 score on the validation dataset (0.55 vs 0.54). While improvements due to data augmentation currently seem marginal, note that we only incorporated 1000 additional images from the unlabeled Syntax dataset derived from the same overarching ARCADE dataset. There are vast quantities of unlabeled XCA angiography data, and being able to access them all for training would likely create both more effective and generalizable models.

This was demonstrated through StenosisSeg-other, a model that incorporated 1000 images from the Danilov et al. [3] dataset as well. While F1 score decreased to 0.47 on the test set likely because of incorporation of non-ARCADE images, the model had a broader range of training images and we hypothesize it will likely perform better in the real world.

All StenosisSeg models significantly outperformed the baseline YOLOv8-seg model, which achieved an F1 score of 0.35 and 0.31 on validation and test sets, respectively.

6.2. Future Work

Due to the surprising result that diversifying our training dataset to images collected from different sources hindered our performance on the ARCADE test set, we propose development of XCA stenosis segmentation benchmarks with images from diverse sources as an avenue of future work. We hypothesize that this will strengthen future model development, as the data would be more diverse and reflective of a clinical setting.

In addition, we would like to contact radiologists to receive their qualitative hypotheses on why our model might be providing incorrect stenosis segmentations in our failure cases. If there are recurring patterns (i.e. artifacts during imaging, common anatomical structures visually similar to stenosis), perhaps we can employ some form of

contrastive loss to capture these features and minimize our error.

Finally, we would like to bolster our data augmentation pipeline further. This could include adding geometric transformations, color space transformations, blurring, or noising. Another notable future improvement could be utilizing "soft" [7] masks instead of binary classification masks, which might provide a less black-and-white approach to segmenting stenosis on XCA images where there is inherent uncertainty.

7. Contributions and Acknowledgment

T.J., A.W., and E.L. contributed equally to all parts of the project including conceptualization, ideation, literature review, methodology development, code implementation, and manuscript preparation. We would like to thank Bohan Wu for his feedback on our ideation and methodology. We would also like to thank Stanford University and the entire CS 231N teaching team for providing funding for cloud computing resources. Pytorch, and specifically the Pytorch Image Models (timm) library, was used in the creation of the training, evaluation, and visualization code for Mask R-CNN development [16].

References

[1] M. Bilal, D. Martinho, R. Sim, A. Qayyum, H. Vohra, M. Caputo, T. Akinosho, S. Abioye, Z. Khan, W. Niaz, et al. Multivessel coronary artery segmentation and stenosis localisation using ensemble learning. *arXiv preprint arXiv:2310.17954*, 2023. 2

[2] A. Chetlen, T. Chan, D. Ballard, A. Frigini, A. H. Hildebrand, S. Kim, J. Brian, E. Krupinski, and D. Ganeshan. Addressing burnout in radiologists. *Academic Radiology*, 26:526–533, 2019. 1

[3] V. Danilov, K. Klyshnikov, O. Gerget, A. Kutikhin, V. Ganyukov, A. Frangi, and E. Ovcharenko. Real-time coronary artery stenosis detection based on modern neural networks. *Scientific Reports*, 11, 2021. 2, 3, 4, 6

[4] T. Du, L. Xie, H. Zhang, X. Liu, X. Wang, D. Chen, Y. Xu, Z. Sun, W. Zhou, L. Song, et al. Training and validation of a deep learning architecture for the automatic analysis of coronary angiography: Automatic recognition of coronary angiography. *EuroIntervention*, 17(1):32, 2021. 2

[5] H. R. Fazlali, N. Karimi, S. R. Soroushmehr, S. Shirani, B. K. Nallamothu, K. R. Ward, S. Samavi, and K. Najarian. Vessel segmentation and catheter detection in x-ray angiograms using superpixels. *Medical & biological engineering & computing*, 56:1515–1530, 2018. 2

[6] R. Gharleghi, N. Chen, A. Sowmya, and S. Beier. Towards automated coronary artery segmentation: A systematic review. *Computer Methods and Programs in Biomedicine*, 225:107015, 2022. 1

[7] C. Gros, A. Lemay, and J. Cohen-Adad. Softseg: Advantages of soft versus binary training for image segmentation. *Medical image analysis*, 71:102038, 2021. 7

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2, 3

[9] W. Huang, L. Huang, Z. Lin, S. Huang, Y. Chi, J. Zhou, J. Zhang, R.-S. Tan, and L. Zhong. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. In *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 608–611. IEEE, 2018. 1

[10] G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics YOLO, Jan. 2023. 2

[11] I. K. Lee, J. Shin, Y.-H. Lee, J. Ku, and H.-W. Kim. Ssass: Semi-supervised approach for stenosis segmentation. *arXiv preprint arXiv:2311.10281*, 2023. 2, 3, 4

[12] F.-F. Li and E. Adeli. Cs231n: Deep learning for computer vision, 2024. Course, Stanford University. 3

[13] Y. Li, Y. Wu, J. He, W. Jiang, J. Wang, Y. Peng, Y. Jia, T. Xiong, K. Jia, Z. Yi, et al. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. *European Radiology*, 32(9):6037–6045, 2022. 1

[14] H. Lin, T. Liu, A. Katsaggelos, and A. Kline. Stenunet: Automatic stenosis detection from x-ray coronary angiography. *arXiv preprint arXiv:2310.14961*, 2023. 2

[15] D. Müller, I. Soto-Rey, and F. Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210, 2022. 4

[16] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. *CoRR*, abs/1912.01703, 2019. 7

[17] S. Pokhrel, S. Bhandari, E. Vazquez, Y. R. Shrestha, and B. Bhattarai. Data augmentation through pseudolabels in automatic region based coronary artery segmentation for disease diagnosis. *arXiv preprint arXiv:2310.05990*, 2023. 2

[18] M. Popov, A. Amanturdieva, N. Zhaksylyk, A. Alkanov, A. Saniyazbekov, T. Aimyshev, E. Ismailov, A. Bulegenov, A. Kuzhukeyev, A. Kulanbayeva, et al. Dataset for automatic region-based coronary artery disease diagnostics using x-ray angiography images. *Scientific Data*, 11(1):20, 2024. 1, 2, 3, 4, 5, 6

[19] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015. 2, 3

[20] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, C. A. Anderson, P. Arora, C. L. Avery, C. M. Baker-Smith, A. Z. Beaton, A. K. Boehme, A. E. Buxton, et al. Heart disease and stroke statistics—2023 update: a report from the american heart association. *Circulation*, 147(8):e93–e621, 2023. 1

[21] H. Zhang, Z. Gao, D. Zhang, W. K. Hau, and H. Zhang. Progressive perception learning for main coronary segmentation

in x-ray angiography. *IEEE Transactions on Medical Imaging*, 42(3):864–879, 2022. 2