Enhancing M-mode Echocardiography Labeling with Instance Segmentation: A Study on Label Assignment Methods

Min Luo

March 29, 2024

Abstract

M-mode echocardiography is a method used to detect defects in the heart by utilizing ultrasound to assess various parameters such as the thickness of cardiac structures, which can reveal abnormalities. Labeling the different parts of the cardiac vasculature for thickness measurements is a laborious and time-consuming task when done manually, thus instance segmentation can assist in this process.

REMEM [6] introduces an instance segmentation model capable of generating masks for different cardiac structures and measuring their thickness, along with proposing a dataset MEIS for this purpose. This paper aims to discuss the accuracy of the instance segmentation model, with a focus on label assignment, and examines various IoU measurement methods to determine the most suitable approach for this type of modeling task.

Label assignment adopts the SimOTA[3] used by YOLOX and the Soft modified by RTMdet [4]. Ultimately, it is found that RTM performs the best and is more suitable when combined with traditional IoU measurements. This

improvement results in a transition of the model from one level of performance to another. The code is on https://github.com/minmin1223/heart.

1 Introduction

Cardiac ultrasound is a primary choice for cardiac assessment. It is a fundamental imaging modality used in the evaluation of cardiovascular diseases, which offers comprehensive information about the structures and functions of the heart. [1]

To visualize distinct structures of the left ventricle (LV) during systolic and diastolic phases, a sequence of period images is recorded using m-mode echocardiography. While m-mode echocardiography offers real-time imaging, the process of manually labeling heart structures is time-consuming. Addressing this challenge, real-time instance segmentation emerges as a pivotal solution for expediting the measurement of heart diseases. The real-time automatic M-mode echocardiography measurements, as introduced in [6], present a method that achieves both accuracy and speed in generating masks for different parts of the left ventricle (LV) and Aortic Valve (AV).

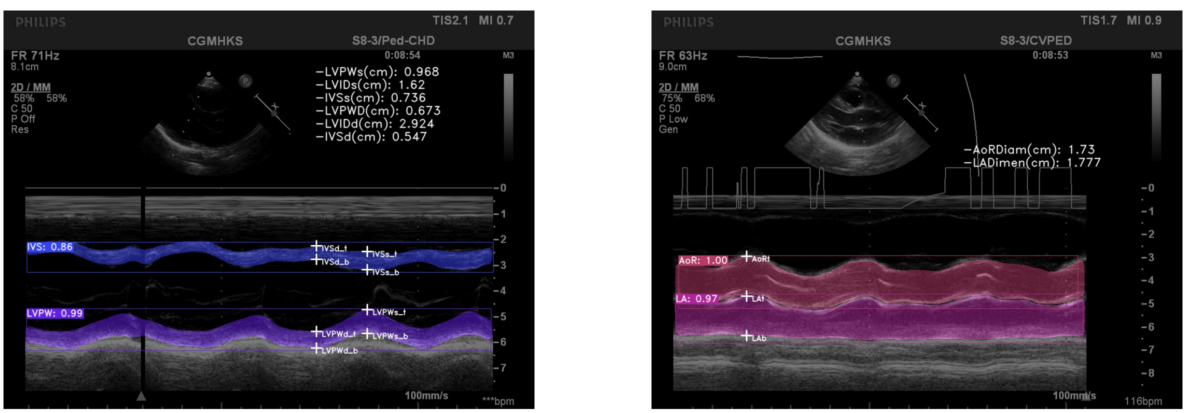

Figure 1: The left image is the instance segmentation annotation on image of Left Ventricle (LV), the right image is the instance segmentation annotation on image of Aortic Valve (AV)

To achieve higher precision, label assignment is a crucial issue in the field of computer vision. ATSS[7] emphasizes the importance of defining positive and negative training samples. OTA [3] aims to address the challenges posed by ambiguous anchors, thereby improving the model's capability to differentiate between distinct objects. The challenge of M-mode echocardiogram images is the indistinguishable boundaries between different structure of heart, therefore, the intent of this work is to resolve this issue by introducing a different approach of label assignment. The contributions of the paper are outlined as follows:

- Enhances the model accuracy by improving the ability to distinguish ambiguous boundaries in cardiac structures.
- Provides an analysis of the maIOU method with various approaches to label assignment.

2 Related Work

2.1 Real-time instance segmentation

Instance segmentation is a computer vision task that involves identifying and delineating individual objects within an image. Unlike semantic segmentation, which classifies each pixel in an image into predefined classes, instance segmentation goes a step further by distinguishing between individual instances of objects. In other words, it aims to provide a unique label for each distinct object within the image and produce a pixel-wise mask for each instance. Real-time performance implies that the segmentation algorithm can provide results in 0.33 sec (¿30 FPS). [2]

The approach of YOLACT [2] breaks instance segmentation into two tasks: generating prototype masks and predicting per-instance mask coefficients. Instance masks are produced by linearly combining prototypes with coefficients, resulting in high-quality masks. Additionally, YOLACT propose Fast NMS, a 12 ms faster replacement for standard NMS with minimal performance impact.

maLYOLACT [5], a Continuation work of YOLACT[2], introduces mask-aware IoU for improving anchor box assignment in instance segmentation training. maIoU incorporates the mask into conventional IoU, providing more accurate supervision to the predicted anchors.

RAMEM[6], a real-time automatic M-mode echocardiography measurement scheme provides an open dataset for instance segmentation, MEIS, abbreviated for M-mode echocardiograms for instance segmentation, proposes panel attention to address. RAMEM contributes by providing MEIS, an M-mode echocardiogram dataset, introducing efficient panel attention for global receptive fields, last of all, implementing AMEM for automatic measurement labeling.

2.2 Label Assignment

In computer vision, label assignment is crucial for tasks such as object detection and segmentation. Proper label assignment is essential for training accurate and effective models, as it forms the basis for learning the relationships between visual features and corresponding object classes. The accuracy of label assignment directly impacts the performance of computer vision algorithms in task, such as distinguishes objects between foreground and background, allowing for further refinement of positional accuracy specifically for foreground objects. . ATSS[7] emphasizes the importance of defining positive and negative training samples. The proposed Adaptive Training Sample Selection (ATSS) automatically selects samples based on object statistics, significantly improving both anchor-free and anchor-based of detectors. The method bridges the performance gap between anchor-based and anchor-free detectors.

ATSS has made significant breakthroughs in the task of label assignment, while OTA proposes improvements to the model to address the issue of simultaneously assigning anchors to two ground truths. OTA [3] introduces a novel approach to label

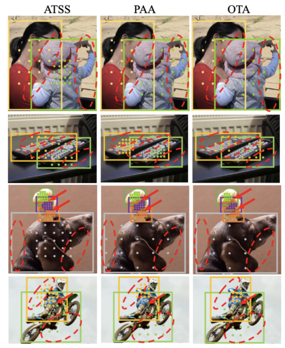

Figure 2: The picture shows anchors with different method of label assignment, OTA solved the problem of ambiguous anchors.

assignment in object detection by formulating it as an Optimal Transport problem. The unit transportation cost is defined as a weighted sum of classification and regression losses between anchors and ground-truth objects. The method particularly address the problem of ambiguous anchors between different ground truth boxes, and performs exceptionally well in crowded scenarios.

RTMDet [4] is a real-time object detector. Its primary contributions include the incorporation of a novel architecture featuring large-kernel depth-wise convolutions and the introduction of soft labels for dynamic label assignment. The label assignment process enhances OTA by utilizing soft labels, thereby improving the alignment between predicted objects and ground truth objects.

3 Method

Similar to OTA, Soft Dynamic Label Assignment [4] employs a cost function to estimate the adaptability of predicted anchors. The key difference is the use of soft labels or IoU Y_{soft}, which replace hard labels, and the incorporation of Binary Cross Entropy instead of Focal Loss in calculating the classification cost.

$$C = \lambda_1 C_{cls} + \lambda_2 C_{reg} + \lambda_3 C_{centers}$$

With classification score S, soft label Y, and Euclidean distance, the classification cost C_{cls} is formulated as below.

$$C_{cls} = BCE(S, Y_{soft})Distance(S, Y_{soft})$$

However, in order to simplify the formulation and improve the efficiency, Binary Cross Entropy is replaced by logarithm as following.

$$C_{cls} = -log(S)Distance(S, Y_{soft})$$

In order to explicitly improve the anchors with lower IoU, logarithm is used in regression cost function.

$$C_{reg} = -log(IoU)$$

For center cost, RTMDet adopts soft center region cost as following. α and β are the default set as 10 and 3.

$$C_{center} = \alpha^{|x_{pred} - x_{gt}| - \beta}$$

4 Experiments

Without incurring additional costs in the experiment, this study modifies the RAMEM [6] model by incorporating soft dynamic label assignment from [4] and employing maIOU as the original model.

Table 1 presents a comparative analysis of three object detection methods: SimOTA (Original OTA), Soft dynamic label assignment, and ATSS (Adaptive Training Sample Selection). The mean Average Precision (mAP) is provided for each method, along with the maIOU metric.

RAMEM employs ATSS with maIOU. Both variants of SimOTA, with and without maIOU, exhibit lower mAP scores compared to the other methods. Conversely, both versions of Soft dynamic label assignment demonstrate higher mAP scores. Particularly noteworthy is that the version of Soft dynamic label assignment without maIOU, utilizing traditional IOU, outperforms its maIOU counterpart. This observation suggests that maIOU might not be the optimal method for soft dynamic label assignment.

SimOTA	Soft dynamic label assignment	ATSS	maIoU	mAP
V				39.8
V			V	41.94
		V		44.485
		V	V	45.51
	V		V	45.58
	V			45.83

Table 1: 3 methods of label assignment with maIoU

5 Conclusion

The version of Soft dynamic label assignment without maIOU, relying on traditional IOU, surpasses its maIOU counterpart in performance. This observation challenges the notion that maIOU is the optimal method for soft dynamic label assignment. Therefore, our results suggest that further exploration and evaluation of alternative label assignment strategies, particularly in the context of instance segmentation on M-mode echocardiogram images, are warranted to enhance model performance and accuracy.

References

[1] Islam Aly, Asad A. Rizvi, Wallisa Roberts, Shehzad Khalid, Mohammad W. Kassem, Sonja Salandy, Maira du Plessis, R. Shane Tubbs, and Marios Loukas. Cardiac ultrasound: An anatomical and clinical review. *Translational Research in Anatomy*, 2020.

[2] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact++ better real-time instance segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:1108–1121, 2019.

[3] Zheng Ge, Songtao Liu, Zeming Li, Osamu Yoshie, and Jian Sun. Ota: Optimal transport assignment for object detection. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 303–312, 2021.

[4] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. Rtmdet: An empirical study of designing real-time object detectors. *ArXiv*, abs/2212.07784, 2022.

[5] Kemal Oksuz, Baris Can Cam, Fehmi Kahraman, Zeynep Sonat Baltaci, Sinan Kalkan, and Emre Akbas. Mask-aware iou for anchor assignment in real-time instance segmentation. *ArXiv*, abs/2110.09734, 2021.

[6] Ching-Hsun Tseng, Shao-Ju Chien, Po-Shen Wang, Shin-Jye Lee, Weiping Hu, Bin Pu, and Xiaojun Zeng. Real-time automatic m-mode echocardiography measurement with panel attention from local-to-global pixels. *ArXiv*, abs/2308.07717, 2023.

[7] Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z. Li. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9756–9765, 2019.