2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-8622-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/BIBM62325.2024.10821757

From Non-Expert to Expert: Recurrent Refined Learning for Medical Image Segmentation

Ruohui Jiang[1, 2†], Changtai Li[1, 2†], Xiaojuan Ban[1, 2, 3*], Shihua Yin[4], Chao Yao[5], Yu Guo[1],
Mohammad S. Obaidat, Life Fellow of IEEE[6, 7*]

[1]Beijing Advanced Innovation Center for Materials Genome Engineering,
University of Science and Technology Beijing, Beijing, China
[2]School of Intelligence Science and Technology, University of Science and Technology Beijing, Beijing, China
[3]Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang, China
[4]Second Affiliated Hospital of Guangxi Medical University, Nanning, China
[5]School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China
[6]King Abdullah II School of Information Technology, University of Jordan, Amman, Jordan
[7]Amity University, Noida, UP, India
Email: jiangruohui32@163.com, lichangtai17@gmail.com, banxj@ustb.edu.cn, shihuayin@126.com,
yaochao@ustb.edu.cn, guoyu@ustb.edu.cn, m.s.obaidat@ieee.org

Abstract—**Precise lesion segmentation is essential in computer-aided diagnosis and treatment. Prevalent deep-learning-based approaches need high-quality annotations to achieve satisfying performance. However, blurring effects and infiltration hamper the accurate delineation of lesions. A common solution is to divide the annotation process into initial annotation by non-specialized personnel and subsequent modification by expert physicians, where the latent correction patterns and non-expert labels are seldom utilized effectively. To explore their helpfulness, we propose ReReNet, a recurrent refined network for lesion segmentation that learns from non-expert to expert. It achieves progressively refined segmentation results through multiple iterations with tailored discrepancy-aware supervision. During training, as the iterative process perceives discrepancies between refining and expert labels, the model gradually grasps the knowledge of turning the barely correct into the clinically accurate. We validate ReReNet's capability on three medical image segmentation (MIS) datasets, including magnetic resonance (MR) and computed tomography (CT) modalities. Comparison results indicate that the proposed approach achieves superior performance by introducing the designed recurrent mechanism and outperforms mainstream methods, demonstrating the effectiveness of mining hidden correction patterns by utilizing non-expert information.**

Index Terms—**Medical imaging, Lesion segmentation, Coarse to fine, Deep learning**

I. Introduction

Medical image segmentation (MIS), identifying and isolating critical regions like lesions, blood vessels, and tissues [1], plays a major role in early diagnosis, staging, treatment planning, and predicting prognosis [2]. Imaging techniques like computed tomography (CT), magnetic resonance (MR), etc., offer non-invasive ways to get detailed information about lesions, including their relationships to surrounding structures [3], [4]. To automate the delineation process requiring time and expertise, numerous techniques are developed and applied to real-world scenarios [5], [6], among which prevailing deep-learning-based ones are recognized to identify complex patterns from massive data [1], [7].

However, training a capable model needs abundant high-quality annotations [8], [9]. For MIS, annotating presents distinct challenges compared to natural scenes. The inherent complexity of medical images requires extensive domain expertise, hindering accurate labeling by untrained individuals [10]. Conversely, annotating in natural scenes for image segmentation or object detection can be delegated to non-experts, enabling rapid dataset curation. The difficulty in recognizing regular targets (buildings, persons, lane lines, etc.) is minimal. In contrast, medical MR and CT imaging necessitate careful judgment and confirmation for seasoned technicians to identify critical regions [11], [12], especially in complex scenarios with blurred boundaries and intermingled tissues.

A prevalent solution in practice for lesion annotation adopts a two-step approach [13], as shown in Fig. 1 (a). Firstly, non-expert annotators with brief training perform coarse labeling, capturing diagnostically relevant information like location and size. Secondly, experts, typically senior physicians, review and correct these coarse annotations. This strategy balances annotation quality with resource utilization, significantly reducing expert workload and overall costs. However, these non-expert labels are typically overlooked during the training of the lesion segmentation model. Indeed, mislabeling occurs more frequently among non-experts due to the uncertainty of blurred boundaries and tumor infiltration. Fig. 1 (b) depicts the discrepancies between non-expert labels and ground truth. We argue that regardless of their accuracy, modeling the correction process from unreliable (non-expert) to reliable (expert) labels can guide the model to explicitly discover refinement patterns and leverage helpful information in non-expert labels, thereby achieving more precise and robust lesion segmentation.

Therefore, we propose ReReNet, a recurrent refined frame-

* Corresponding author: banxj@ustb.edu.cn, m.s.obaidat@ieee.org.

† These authors have contributed equally to this work.

This paper is sponsored by the National Key Research and Development Program of China (2022ZD0118001) and the National Natural Science Foundation of China under Grant U22A2022, 62332017 and 62303043.

979-8-3503-8622-6/24/$31.00 ©2024 IEEE

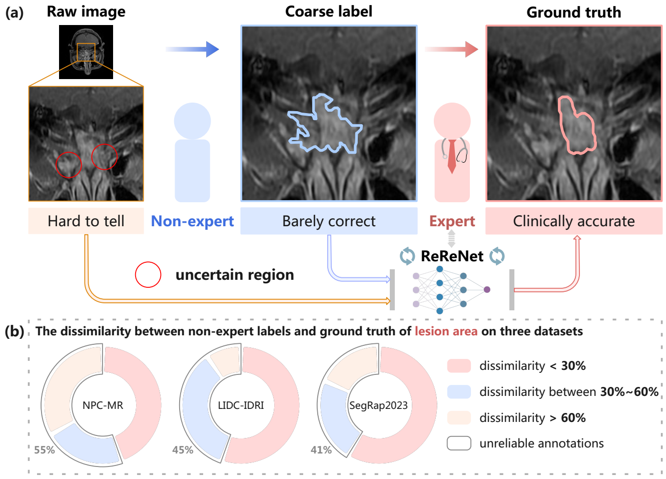

Fig. 1. The motivation of ReReNet. (a) A common way to generate ground truth is for a non-expert to delineate the lesion coarsely, followed by refinement by an expert. ReReNet aims to mimic the refining process by recurrently pushing the model to output the expert-like mask. (b) On the three datasets, nearly half of the lesions have a dissimilarity of more than 30% in lesion area between coarse labels and ground truth.

work for lesion segmentation. To the best of our knowledge, this is the first study of utilizing non-expert labels to assist in MIS tasks. Specifically, we treat the non-expert label as the initial coarse label. The output mask feeds into the model recurrently as the coarse label of the next refinement stage. Those iterative processes continually enhance the accuracy of the segmentation. Further, to address the issue countered when naively learning with the blunt repetition that the performance unexpectedly decreases instead of increases, we design a discrepancy-aware optimization strategy to calculate the Discrepancy loss that explicitly considers label differences and guides the model toward better learning from coarse labels.

We assess the performance and validate the effectiveness of ReReNet on three datasets, NPC-MR (in-house), LIDC-IDRI [14], and SegRap2023 [15]. Compared with other mainstream MIS methods, ReReNet performs better by mimicking the correction process conducted by expert physicians from gradually refined coarse labels. Taking a naive U-Net [16] as the cornerstone, the modified version (ReReNet) attains a higher DSC by 4.36%, 1.17%, and 1.11%, respectively. Furthermore, we conduct segmentation experiments at different label ratios during training on the NPC-MR dataset. With only 10% of labels, ReReNet achieves the performance of the baseline model trained on 50% of labels. This substantial reduction in reliance on high-standard annotations highlights ReReNet's potential in MIS.

The main contributions of our work are summarized below:

- We propose ReReNet, a novel learning framework to harness non-expert labels that may have been previously overlooked and to refine segmentation results recurrently.
- We develop a discrepancy-aware optimization strategy, perceiving the discrepancies with expert labels as the process iterates. This strategy of calculating the loss guides the network model to focus on these discrepancies.
- We conduct extensive experiments on three lesion segmentation datasets, of which the numerical and visual results illustrate that the proposed ReReNet delivers higher segmentation accuracy and alleviates the label demand compared to typical methods. Ablation studies also prove the efficacy of the recurrent mechanism and the discrepancy-aware supervision.

II. Related Work

In the past decade, the boosted computational capacity encouraged the development of numerous deep learning-based automatic MIS algorithms. After fully convolutional networks (FCNs) [17], the emergence of the simple yet effective U-Net [16] established the U-shaped architecture as a de facto model in medical image analysis in either 2D [18], [19] or 3D [20], [21]. Recently, transformer has been introduced into MIS, such as UNETR [22], TransUNet [23], and Swin-Unet [24], etc. Several essential phenomenons in Medical images that raise the difficulty of accurately segmenting the tumor lesion are blurred lesion boundaries and noises [10]. To overcome the issue, some methods adopt the coarse-to-fine paradigm [25]–[28]. The first stage generates an approximate region, and the second stage refines this region using more sophisticated models. CFU-Net [29] introduces an additional decoding path at the feature level, using the coarse part to guide the decoding of the refined part. [30] utilizes semantic information from feature maps of different layers to refine the boundaries of coarse masks, thereby enhancing segmentation performance.

However, few of the above methods have attempted to leverage non-expert annotations to surmount the uncertainty caused by blurred and intermingled lesion regions. Using coarse labels, we innovatively formulate the recurrent refined network to model the coarse-to-fine pattern from non-expert labels to expert labels, thereby realizing accurate lesion segmentation.

III. Method

This paper presents a novel learning framework to harness abundant coarse label information previously underutilized and introduces a discrepancy-aware optimization strategy to enhance this learning paradigm. We first formulate the problem of segmenting the lesion or organ region and clarify the mathematical notations. Next, we delve into the specific concept and methodology of the framework. Based on that, we describe the dedicated optimization strategy tailored to this scenario. Fig. 2 demonstrates the detail of the proposed ReReNet and the Discrepancy loss.

A. ReReNet Framework

Inspired by the coarse-to-fine thought, we design a recurrent and progressive learning architecture that utilizes coarse labels to assist in segmentation. Non-expert labels are treated as the initial coarse labels. The output masks are fed into the model recurrently as the coarse labels of the next refinement stage, continually enhancing the segmented results.

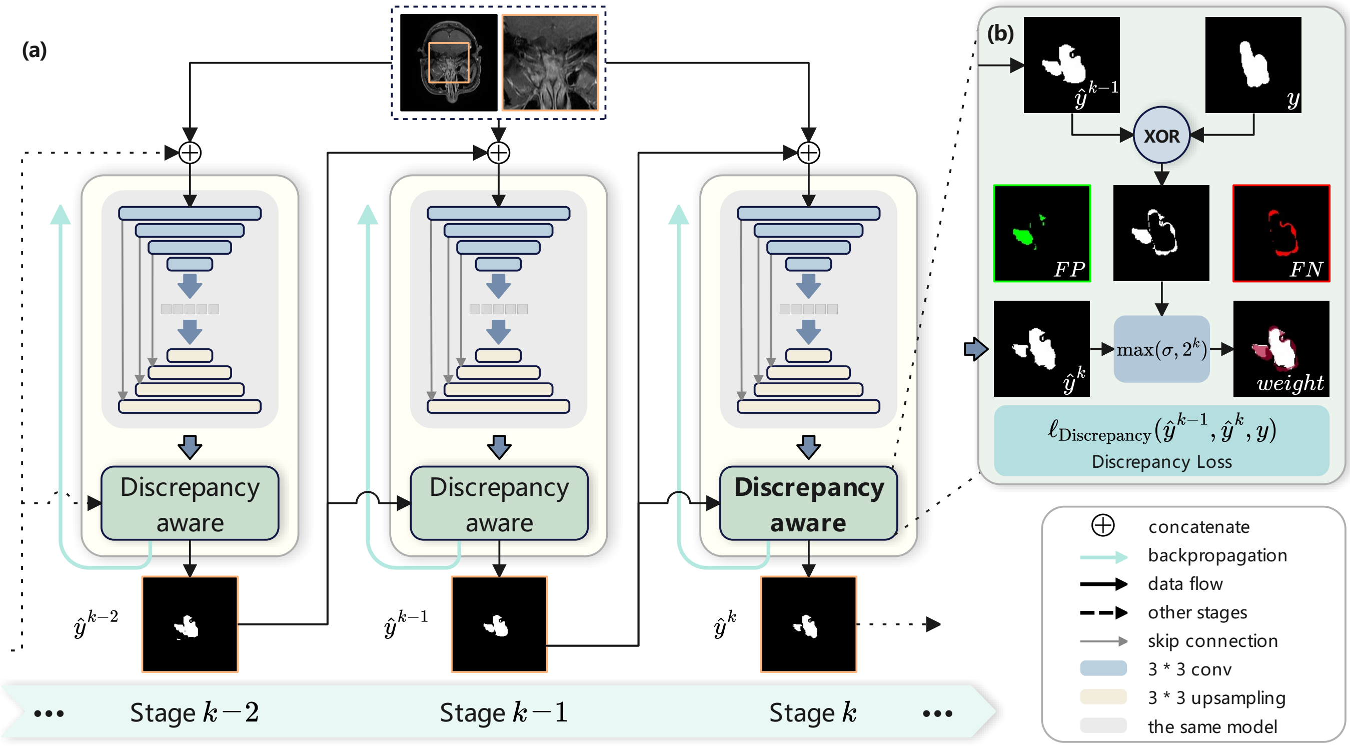

Fig. 2. The scheme of the proposed ReReNet. (a) The overall data flow of the recurrent process from stage $k-2$ to stage k is depicted. The output mask of one stage is reused as the input of the next stage. (b) The discrepancy-aware optimization strategy is demonstrated. The discrepancy is obtained by the coarse label and ground truth to weight the loss.

Considering the input space $\mathcal{X}$ (i.e., the image space), each image comprises a set of pixels N, where $N = H \times W$. In a standard semantic segmentation setup, given an image $x \in \mathcal{X}$, our objective is to learn a mapping that assigns a label $y_i \in \mathcal{Y}$ to each pixel x_i, representing its semantic category. The mapping is achieved by a learnable encoder-decoder model f_θ, predicting the probability of mapping from the image space $\mathcal{X}$ to the label space $\mathcal{Y}$. The output segmentation mask is $\hat{y} = \{\arg\max_{c \in \mathcal{Y}} p_i^c\}_{i=1}^N$, where p_i^c is the model's predicted probability of pixel x_i belonging to category c. Hence, $\hat{y} = f_\theta(x)$.

In the training and inference phase of ReReNet, the final segmentation mask is obtained through multiple stages, as shown in Fig. 2 (a). We divide this process into $\mathcal{K}$ stages. During training, at each stage k, the output mask $\hat{y}^{k-1}$ from the previous stage is input into the network, aiding in producing a more refined segmentation outcome, which is denoted as:

$$\hat{y}^k = f_\theta^{k-1}(x, \hat{y}^{k-1}). \tag{1}$$

The parameters of the model f_θ^{k-1} are updated by calculating the loss between $\hat{y}^k$ and y to obtain f_θ^k for the subsequent stage. Upon the completion of training, we have the final model f_θ^*. It is noted that the recurrent mechanism is also employed in inference, where the mask from the previous stage concatenates with the raw image to input into the model. After iterating $\mathcal{K}$ times, the whole process is finished and generates the final segmentation result $\hat{y}^k = f_\theta^*(x, \hat{y}^{k-1})$.

For a given image x, during the segmentation process, the image x and the initial coarse label (non-expert label) $\hat{y}^0$ are concatenated and merged along the feature channel to serve as the input into the model. ReReNet outputs the segmentation mask $\hat{y}^1$. Subsequently, the discrepancy between $\hat{y}^1$ and the ground truth y is obtained for the loss calculation, and the network's parameters are updated through gradient descent. This step is defined as stage 1, whose output $\hat{y}^1$ is utilized as the coarse label for the next stage. The process is repeated $\mathcal{K}$ times to complete the training of one batch, and the inference process follows the same procedure.

B. Discrepancy-Aware Optimization Strategy

Upon the ReReNet, we also propose a discrepancy-aware optimization strategy to optimize the model's training. Mainstream approaches commonly use the Dice loss for network optimization. Due to the varying sizes of target regions, it is more robust against class imbalance issues and may offer greater stability during training. The general form of the Dice loss is as follows:

$$\ell_{\text{Dice}}(\hat{y}, y) = 1 - \frac{2\sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i}. \tag{2}$$

However, solely using the Dice loss overlooks the valuable information on the coarse label. To effectively utilize this information, we introduce a discrepancy-aware optimization strategy, i.e., a loss function, enhancing the network's perception of the discrepancy.

The following describes the difference-aware optimization strategy and its implementation. As shown in the discrepancy-aware module in Fig. 2 (b). As understood from the ReReNet framework paradigm, the inputs to the network are $x, \hat{y}^{k-1}$, where x represents the original input image, and $\hat{y}^{k-1}$ is the coarse label from the previous stage. The network output is $\hat{y}^{k}$, the current stage's output mask. The discrepancy is defined as $\text{discrepancy} = \text{XOR}(\hat{y}^{k-1}, y)$, representing the portions incorrectly classified by the model in the previous stage, containing the false positive (FP) and false negative (FN). For these mis-segmentation regions, a corresponding weight map is built for parameters' optimization, manifested as a loss function. The weight w is used for calculating the loss, where λ is the weight factor.

$$w = \text{XOR}(\hat{y}^{k-1}, y) * \lambda + 1, \tag{3}$$

based on the weight map, the Discrepancy loss can be derived as:

$$\begin{aligned}\ell_{\text{Discrepancy}}(\hat{y}^{k-1}, \hat{y}^{k}, y) = -\sum_{i}^{N}(w_i * (y_i \cdot log(\hat{y}_i^k) \\ +(1-y_i) \cdot log(1-\hat{y}_i^k))).\end{aligned} \tag{4}$$

Incorporating both ℓ_{Dice} and $\ell_{\text{Discrepancy}}$, the overall network optimization loss for the ReReNet can be expressed as:

$$\begin{aligned}\ell(\hat{y}^{k-1}, \hat{y}^{k}, y) = \alpha \cdot \ell_{\text{Dice}}(\hat{y}, y) \\ +(1-\alpha) \cdot \ell_{\text{Discrepancy}}(\hat{y}^{k-1}, \hat{y}^{k}, y),\end{aligned} \tag{5}$$

where α is the hyperparameter that balances the two losses. To address the unexpected drop in performance observed with naive cyclic learning concerning the different stages of k, we assign varying weights to the different stages. This intuitively reflects that, with the progression of recurrent iterations, the model incrementally intensifies the penalty for errors, expecting better results, as demonstrated below:

$$\ell(\hat{y}^{k-1}, \hat{y}^{k}, y)^k = \max(\sigma, 2^k) * \ell(\hat{y}^{k-1}, \hat{y}^{k}, y), \tag{6}$$

where σ is the initial factor. We optimize the entire model across various stages based on $\ell(\hat{y}^{k-1}, \hat{y}^{k}, y)^k$. Experiments validate the efficacy of our optimization strategy.

IV. Experiments

A. Datasets

In this study, we evaluate the performance of our proposed method on three datasets: NPC-MR, LIDC-IDRI, and SegRap2023.

NPC-MR comprises MR images from 362 nasopharyngeal carcinoma (NPC) patients undergoing treatment. 10 annotators without expertise delineated lesions coarsely, which five experienced medical professionals modified to guarantee clinically accurate labeling, serving as the ground truth. The average dissimilarity, i.e., DSC between these non-expert and expert label pairs, is 68.5%, which is lower than the performance of plain models trained with expert labels. **LIDC-IDRI** [14] includes lung nodule CT scans from 1,018 cases. We selected samples annotated by at least four physicians and cropped them to 256×256 pixels. For a set of multi-observer labels, pixels with an agreement rate exceeding 50% were considered expert ones; the label with the largest deviation from the ground truth was designated as the non-expert label. The average DSC between non-expert and expert pairs is 66.61%. **SegRap2023** [15] contains CT scans with gross target volumes of the nasopharynx (GTVnx) from 200 NPC patients. Notably, simulative non-expert labels are obtained by applying a random perspective transformation to the ground truth. The average DSC between non-expert and expert pairs is 60.63%.

B. Implementation Details

We treat the expert labels as ground truth and the non-expert labels as the initial coarse labels for ReReNet. To ensure a similar distribution of grayscale values in each image, we normalize them using the mean and standard deviation of the entire training set, applying the z-score standardization method.

Other models, except the ReReNet model, take single medical images as input because these models are not initially designed to incorporate additional information and employ the Dice loss for training. The ReReNet model is trained using medical images and coarse labels. In the initial 20 epochs of training, only the Dice loss is used for optimization to avoid the potential instability of introducing the Discrepancy loss in the early learning phase. Models are trained on the training set for 100 epochs, and the best-performing model on the validation set is selected for final metric evaluation on the test set. The Adam optimizer is used to update the parameters. The initial learning rate is set to 1e-4, and the weight decay is set to 1e-5.

For all experiments, hyperparameters were configured: $\lambda = 10$ in (3), $\alpha = 0.8$ in (5), and $\sigma = 1$ in (6). All experiments are carried out on a single NVIDIA GeForce RTX 4090 GPU for training and testing. The operating system is Ubuntu 20.04 LTS, with the PyTorch version 1.13.1.

C. Experimental Results

Multiple evaluation metrics are employed in the experimental section to evaluate model performance comprehensively. The Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are the primary metrics for assessing segmentation accuracy. They provide insights into the overlap between predicted and ground-truth segmentations. To assess boundary differences, the average symmetric surface distance (ASSD) and 95th Percentile Hausdorff Distance (HD95) are used. Precision measures the accuracy of positive case predictions, while Recall assesses the model's ability to identify all relevant instances.

1) Segmentation Performance: We assess ReReNet's performance by comparing it with mainstream segmentation models on the NPC-MR, LIDC-IDRI, and SegRap2023 datasets. Of note, unlike semi-supervised learning, which generates

TABLE I

Performance evaluation of ReReNet ($\mathcal{K} = 5$) compared with other typical segmentation models whose implementations are publicly available. Bold marks the best result and underline marks the second best.

Dataset	Method	DSC (%)	IoU (%)	ASSD ↓	HD95 ↓	Precision (%)	Recall (%)
NPC-MR	U-Net	72.19	59.29	5.88	15.70	75.84	74.43
	Attention U-Net	73.18	60.85	7.22	18.35	73.95	78.43
	DeepLabv3	69.90	56.23	6.33	**15.03 (-0.67)**	73.24	72.08
	TransUNet	68.67	55.58	7.60	20.81	68.93	74.43
	Swin-Unet	67.23	53.12	7.08	19.12	69.86	70.97
	ReReNet(ours)	**77.54 (+4.36)**	**66.22 (+5.37)**	**5.36 (-0.52)**	16.39	**78.71 (+2.87)**	**81.39 (+2.96)**
LIDC-IDRI [14]	U-Net	81.28	71.17	3.09	7.05	82.72	84.20
	Attention U-Net	78.63	68.11	4.02	10.75	80.22	82.04
	DeepLabv3	41.15	33.75	51.97	103.55	42.01	43.07
	TransUNet	82.38	71.40	2.18	6.49	82.45	85.16
	Swin-Unet	78.46	67.49	5.79	12.78	79.76	81.04
	ReReNet(ours)	**83.55 (+1.17)**	**73.40 (+2.00)**	**1.13 (-1.05)**	**2.60 (-3.89)**	**84.58 (+1.86)**	**88.39 (+3.23)**
SegRap2023 [15]	U-Net	75.45	62.15	3.86	6.89	75.40	81.62
	Attention U-Net	77.34	64.88	3.51	6.15	76.28	83.31
	DeepLabv3	68.92	54.88	4.94	8.67	70.37	74.84
	TransUNet	74.83	62.00	5.96	9.46	71.35	83.76
	Swin-Unet	67.37	53.72	10.47	20.35	67.30	73.94
	ReReNet(ours)	**78.45 (+1.11)**	**65.96 (+1.08)**	**3.34 (-0.17)**	**5.03 (-1.12)**	**77.28 (+1.00)**	**85.52 (+1.76)**

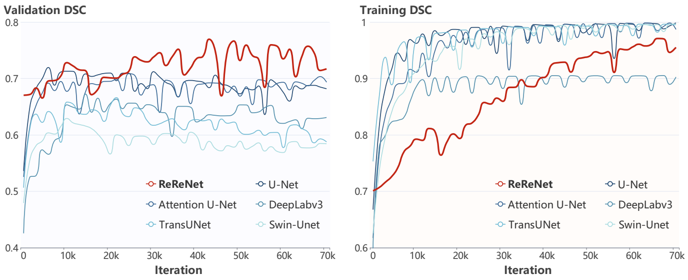

Fig. 3. Trend of the DSC with training iteration on the NPC-MR dataset.

pseudo-labels for unlabeled samples and utilizes them with certain strategies to improve performance [31], our solution is fully supervised; the non-expert labels are provided by non-professionals and have one-to-one corresponding ground truth, which are processed by humans. Therefore, we select representative supervised segmentation architectures for ReReNet to compare with. We choose U-Net as the cornerstone model for our study due to its widespread use and well-established performance in MIS tasks. We also conduct comparative experiments with prominent architectures, including Attention U-Net [18], DeepLabv3 [32], and Transformer-based models like TransUNet [23] and Swin-Unet [24].

As illustrated in Table I, ReReNet demonstrates significantly superior performance to other models. Specifically, on the NPC-MR, ReReNet achieves a 4.36% and 5.37% improvement in the DSC and IoU metrics, respectively, compared to the second-best method. This indicates its relative excellence in segmentation accuracy over other models. Besides, ReReNet achieves a notable increase of 2.87% and 2.96% in Precision and Recall. It is particularly noteworthy that DeepLabv3 has certain advantages in terms of HD95, with a 0.67% reduction. In addition, on the LIDC-IDRI dataset, ReReNet achieves improvements of 1.17% in DSC and 2.00% in IoU compared to other models, demonstrating its superior performance. Finally, ReReNet also shows significant enhancements at SegRap2023. The numerical comparison refers to Table I.

Observing the DSC trend of each model during training, Fig. 3 illustrates the variation in DSC on the validation set during the training process using the NPC-MR dataset. As the training progresses, other methods gradually exhibit signs of overfitting after 10,000 iterations. However, due to the introduction of discrepancy-aware supervision, ReReNet demonstrates enhanced training effectiveness, particularly after 30,000 iterations.

Fig. 4 presents a visual qualitative comparison of the binary masks generated by ReReNet and other models. ReReNet achieves superior segmentation results. For instance, in Fig. 4 (b), where the left side lesions were either incorrectly identified or completely unrecognized by other models, ReReNet accurately recognizes them. Moreover, according to the guide of the non-expert label, ReReNet achieves more accurate segmentation of the right-side lesions.

Overall, the experimental results demonstrate ReReNet's prominent performance in the MIS task, particularly its significant improvements under comprehensive evaluation metrics. These outcomes and findings suggest that the proposed ReReNet can effectively utilize coarse labels to enhance segmentation performance.

2) Performance at Different Label Ratios: To assess the performance of ReReNet with varying ratios of data, we divide the NPC-MR dataset into five different label quantity levels: 10%, 30%, 50%, 70%, and 100%. We compare the segmentation results of ReReNet and the baseline model across these varied label ratios. As illustrated in Table II.

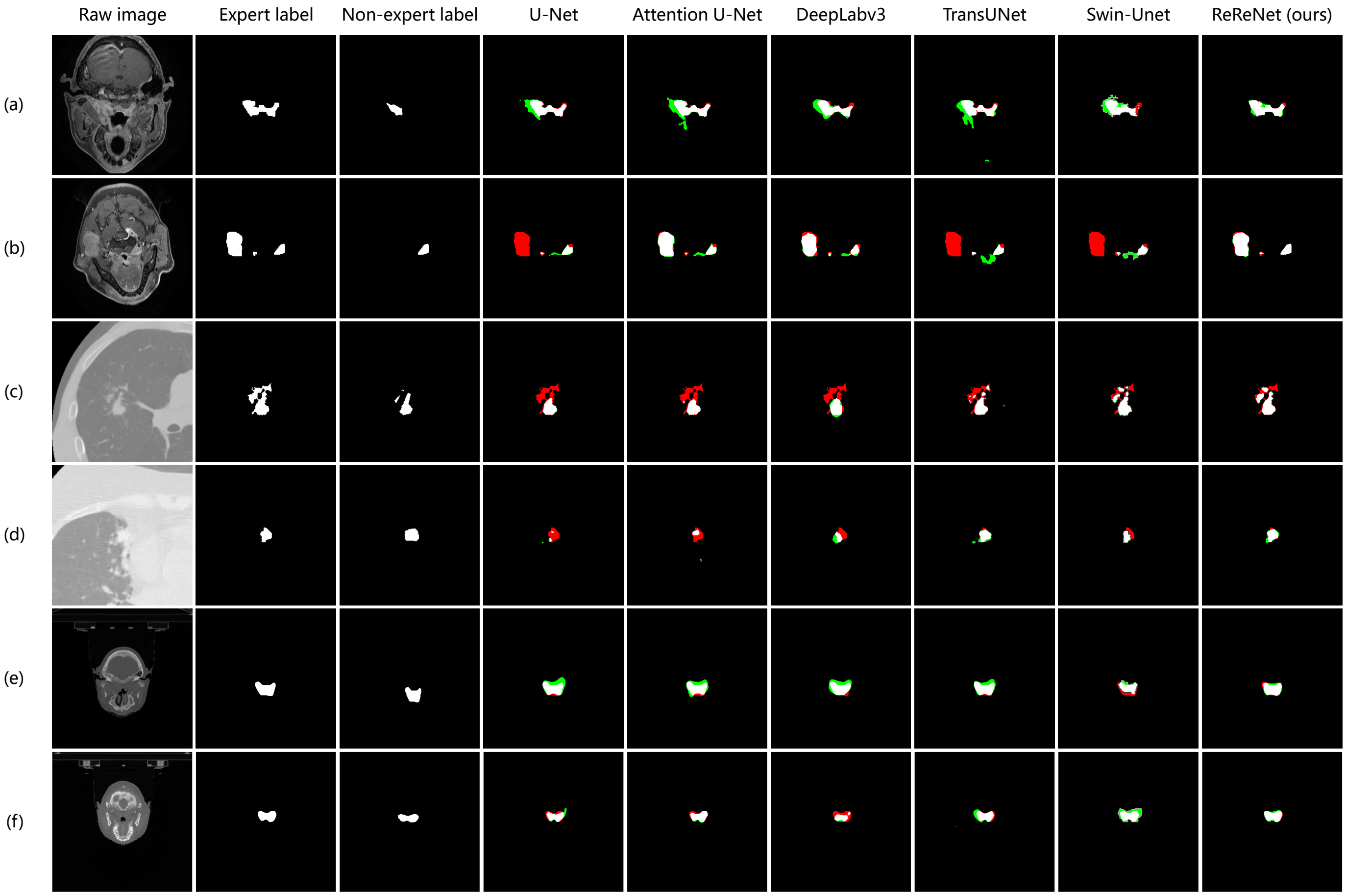

Fig. 4. Qualitative results of ReReNet comparing other approaches. (a) and (b) belong to NPC-MR, (c) and (d) belong to LIDC-IDRI, and (e) and (f) belong to SegRap2023. From U-Net to ReReNet, white denotes correct segmentation, red denotes under-segmentation, and green denotes over-segmentation. Best viewed in color. It can be shown that the segmentation masks output by ReReNet present less over- and under-segmentation.

Our results demonstrate that, as the ratio of labels increases, the model's performance metrics consistently improve. This finding confirms the well-established principle that larger training datasets enhance model performance. Furthermore, it is observed that the accuracy of the baseline model improves more significantly when the label ratio increases from 10% to 50%, while the improvement from 50% to 100% shows diminishing marginal returns. This raises a common trade-off between improving marginal accuracy and the effort of acquiring more labels.

In our experiments, ReReNet significantly reduces the model's reliance on labels, achieving accurate segmentation with fewer labels. ReReNet mostly outperforms the baseline segmentation model across various label ratios. Remarkably, with only 10% labels, ReReNet achieves a better performance (IoU 56.39%) than that of the baseline model trained on 50% labels (IoU 56.01%). On the other hand, by using 70% labels, ReReNet (DSC 74.54%) significantly surpassed the performance of the baseline model with 100% labels (DSC 72.19%).

This demonstrates that ReReNet can achieve outstanding segmentation results even with limited label resources, providing valuable insights for more efficient use of scarce labels. It offers a practical solution to the problem of high-cost labeling in medical imaging, proving the potential of ReReNet for medical image analysis.

3) Visualization of Intermediate Output: Furthermore, as shown in Fig. 5, we visualize the output mask of each step during the iterative inference of ReReNet ($\mathcal{K} = 5$). The figure shows that in the first step, the model's output segmentation image contains many inaccurate regions. However, as it iterates, the model gradually corrects these error regions and finally generates more accurate segmentation results. These visualization results demonstrate ReReNet's ability to gradually eliminate segmentation errors by continuously updating the results in the iterative process and refining its segmentation output with iterations.

D. Ablation Studies

1) Effectiveness of Components of ReReNet: As shown in Table III, on the NPC-MR dataset, we first illustrate the unreliability of non-expert labels. Supervised by only non-expert labels, the trained UNet performs worse than those supervised by expert labels, as anticipated. Then, we verify the

TABLE II

PERFORMANCE GAINS OF ReReNET COMPARED WITH THE BASELINE MODEL AT DIFFERENT LABEL RATIOS ON NPC-MR DATASET.

Metric	Method	Label ratio					Average
		10%	30%	50%	70%	100%	
DSC (%)	U-Net ReReNet	60.34 **68.17 (+7.83)**	65.72 **69.81 (+4.09)**	69.41 **70.23 (+0.82)**	71.21 **74.54 (+3.33)**	72.19 **77.54 (+5.35)**	67.77 **72.06 (+4.29)**
IoU (%)	U-Net ReReNet	46.65 **56.39 (+9.74)**	51.64 **57.83 (+6.19)**	56.01 **58.18 (+2.17)**	58.20 **62.23 (+4.03)**	59.29 **66.22 (+6.93)**	54.35 **60.17 (+5.82)**
HD95 ↓	U-Net ReReNet	43.22 **27.49 (-15.73)**	28.00 **25.56 (-2.44)**	**19.30 (-6.63)** 25.93	**15.38 (-1.74)** 17.12	**15.70 (-0.69)** 16.39	24.32 **22.50 (-1.82)**
Precision (%)	U-Net ReReNet	57.40 **76.08 (+18.68)**	61.20 **73.44 (+12.24)**	73.85 **74.18 (+0.33)**	74.18 **74.56 (+0.38)**	75.84 **78.71 (+2.87)**	68.50 **75.39 (+6.89)**

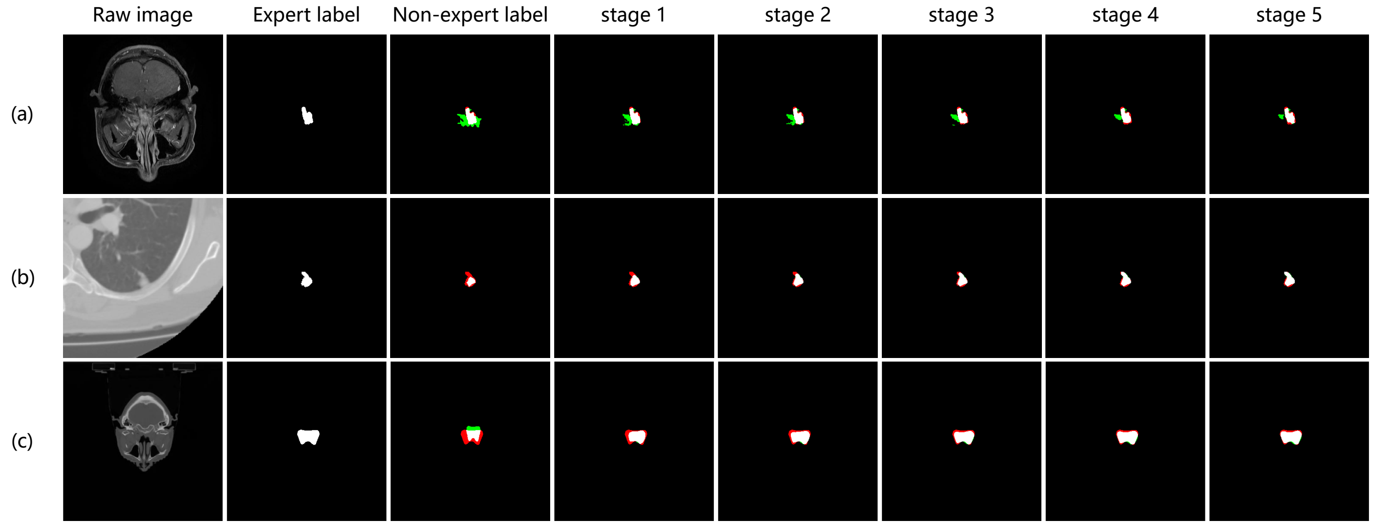

Fig. 5. Intermediate output masks of ReReNet from stage 1 to 5 when $\mathcal{K} = 5$, white denotes correct segmentation, red denotes under-segmentation, and green denotes over-segmentation. Best viewed in color. As the refinement progresses, the segmented masks gradually improved.

TABLE III

ABLATION STUDY OF INTRODUCING NON-EXPERT LABELS, RECURRENT MECHANISM, AND DISCREPANCY-AWARE OPTIMIZATION STRATEGY.

Non-Expert	Expert	Recurrent	Discrepancy	DSC (%)
✓	✗	✗	✗	65.60 (-6.59)
✗	✓	✗	✗	72.19
✓	✓	✓	✗	76.71 (+4.52)
✓	✓	✓	✓	**77.54 (+5.35)**

TABLE IV

ABLATION STUDY OF THE HYPERPARAMETER $\mathcal{K}$.

Total stages	$\mathcal{K} = 1$	$\mathcal{K} = 3$	$\mathcal{K} = 5$	$\mathcal{K} = 7$	$\mathcal{K} = 9$
DSC (%)	77.15	75.36	**77.54**	72.28	72.27

effectiveness of the recurrent mechanism and the discrepancy-aware optimization strategy. ReReNet ($\mathcal{K} = 5$) equipped solely with the recurrent framework achieves a DSC of 76.71%, exceeding the baseline by 4.52%. Incorporating the discrepancy-aware module further boosts DSC by 0.83%, culminating in 77.54%. This improvement underscores the efficacy of the recurrent mechanism and the discrepancy-aware module.

2) Impact of Different Total Stages: We conduct an ablation study on the hyperparameter $\mathcal{K}$ (total stages of ReReNet) on MPC-MR, with results displayed in Table IV. At $\mathcal{K} = 1$, the model achieves favorable results, with a DSC of 77.15%. This validates the effective enhancement of model performance by incorporating coarse labels. As $\mathcal{K}$ increases, there is an initial decline in model performance, which then peaks at $\mathcal{K} = 5$ with a DSC of 77.54%, followed by a substantial decrease as $\mathcal{K}$ continues to rise.

We surmise that when $\mathcal{K} = 1$, a single inference process simplifies the task and potentially leads to superior performance. Increasing $\mathcal{K}$ beyond 1 adds complexity, as the model must correct accumulated errors within the loop. With low $\mathcal{K}$ values, the model struggles to capture the full task distribution. Conversely, a high $\mathcal{K}$ allows the model to learn the task paradigm gradually, but excessive iterations lead to error accumulation that hinders performance. Remarkably, $\mathcal{K} = 5$ appears to strike a balance between model complexity and error accumulation, resulting in the observed optimal performance.

V. Conclusion

In summary, we propose the Recurrent Refined Network, a novel framework that leverages non-expert labels to model the refining process by experts and consequently achieves progressive refinement of segmentation results. Besides, ReReNet applies a discrepancy-aware optimization strategy tailored to coarse-label inputs. Evaluated on three datasets, ReReNet delivers superior performance in lesion segmentation, surpassing the typical architectures. It validates the effectiveness of utilizing non-expert information to identify hidden correction patterns, demonstrating significant potential in the medical image segmentation field.

Several limitations are worth further investigation. Non-expert labels limit the setup of the initial coarse labels in the inference stage. Outputs from rule-based algorithms could be an alternative to suit different application scenarios. Furthermore, our analysis is currently conducted with 2D images. Counterpart 3D architectures should be implemented to cater to the nature of volumetric data in medical imaging.

References

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[2] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and et al., "Medical image segmentation using deep learning: A survey," *IET Image Processing*, vol. 16, no. 5, pp. 1243–1267, 2022.

[3] Z. Wang, M. Fang, J. Zhang, L. Tang, L. Zhong, and et al., "Radiomics and deep learning in nasopharyngeal carcinoma: A review," *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 118–135, 2024.

[4] Z. Lu, Y. Liu, M. Jin, X. Luo, H. Yue, and et al., "Virtual-scanning light-field microscopy for robust snapshot high-resolution volumetric imaging," *Nature Methods*, vol. 20, no. 5, pp. 735–746, 2023.

[5] K. Cao, Y. Xia, J. Yao, X. Han, L. Lambert, and et al., "Large-scale pancreatic cancer detection via non-contrast ct and deep learning," *Nature Medicine*, vol. 29, no. 12, pp. 3033–3043, 2023.

[6] Y. Liu, X. Yuan, X. Jiang, P. Wang, J. Kou, and et al., "Dilated adversarial u-net network for automatic gross tumor volume segmentation of nasopharyngeal carcinoma," *Applied Soft Computing*, vol. 111, p. 107722, 2021.

[7] Y. Zhong, C. Cai, T. Chen, H. Gui, J. Deng, and et al., "PET/CT based cross-modal deep learning signature to predict occult nodal metastasis in lung cancer," *Nature Communications*, vol. 14, no. 1, p. 7513, 2023.

[8] X. Li, W. Tan, P. Liu, Q. Zhou, and J. Yang, "Classification of covid-19 chest ct images based on ensemble deep learning," *Journal of Healthcare Engineering*, vol. 2021, no. 1, 2021.

[9] C. Li, R. Jiang, S. Yin, J. Yang, and X. Ban, "Self-supervised rotation learning for 3d segmentation on nasopharyngeal carcinoma mri images," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Istanbul, 2023, pp. 3529–3534.

[10] Y. Li, T. Dan, H. Li, J. Chen, H. Peng, and et al., "Npcnet: Jointly segment primary nasopharyngeal carcinoma tumors and metastatic lymph nodes in mr images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1639–1650, 2022.

[11] F. Chen, H. Han, P. Wan, L. Chen, W. Kong, and et al., "Do as sonographers think: Contrast-enhanced ultrasound for thyroid nodules diagnosis via microvascular infiltrative awareness," *IEEE Transactions on Medical Imaging*, p. 1, 2024.

[12] Y. Zhang, R. Xi, L. Zeng, D. Towey, R. Bai, and et al., "Structural priors guided network for the corneal endothelial cell segmentation," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 309–320, 2024.

[13] R. Raumanns, G. Schouten, M. Joosten, J. Pluim, and V. Cheplygina, "Enhance (enriching health data by annotations of crowd and experts): A case study for skin lesion classification," *Journal of Machine Learning for Biomedical Imaging*, vol. 1, pp. 1–26, 2021.

[14] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, and et al., "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[15] X. Luo, J. Fu, Y. Zhong, S. Liu, B. Han, and et al., "Segrap2023: A benchmark of organs-at-risk and gross tumor volume segmentation for radiotherapy planning of nasopharyngeal carcinoma," *arXiv preprint arXiv:2312.09576*, 2023.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Munich, 2015, pp. 234–241.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 2015, pp. 3431–3440.

[18] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, and et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[19] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[20] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Athens, 2016, pp. 424–432.

[21] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, 2016, pp. 565–571.

[22] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, and et al., "Unetr: Transformers for 3d medical image segmentation," in *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV*, Waikoloa, 2022, pp. 1748–1758.

[23] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, and et al., "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[24] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, and et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European Conference on Computer Vision (ECCV)*, Tel-Aviv, 2022, pp. 205–218.

[25] L. Xie, Q. Yu, Y. Zhou, Y. Wang, E. K. Fishman, and et al., "Recurrent saliency transformation network for tiny target segmentation in abdominal ct scans," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 514–525, 2020.

[26] Q. Ma, C. Zu, X. Wu, J. Zhou, and Y. Wang, "Coarse-to-fine segmentation of organs at risk in nasopharyngeal carcinoma radiotherapy," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Virtual, 2021, pp. 358–368.

[27] G. Liu, Y. Jiang, D. Liu, B. Chang, L. Ru, and M. Li, "A coarse-to-fine segmentation frame for polyp segmentation via deep and classification features," *Expert Systems with Applications*, vol. 214, p. 118975, 2023.

[28] S. He, Y. Ji, Y. Zhang, A. Zeng, D. Pan, J. Lin, and X. Zhang, "Cfnet: A coarse-to-fine framework for coronary artery segmentation," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Singapore, 2023, pp. 431–442.

[29] H. Yin and Y. Shao, "CFU-Net: A coarse–fine u-net with multilevel attention for medical image segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[30] Y. Li, H. Peng, T. Dan, Y. Hu, G. Tao, and H. Cai, "Coarse-to-fine nasopharyngeal carcinoma segmentation in mri via multi-stage rendering," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Virtual, 2020, pp. 623–628.

[31] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Shenzhen, 2019, pp. 605–613.

[32] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *European Conference on Computer Vision (ECCV)*, Munich, 2018, pp. 801–818.