Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(2): 1617–1636.
DOI: 10.3934/mbe.2023074
Received: 09 August 2022
Revised: 09 October 2022
Accepted: 13 October 2022
Published: 03 November 2022

Research article

Ultrasound carotid plaque segmentation via image reconstruction-based self-supervised learning with limited training labels

Ran Zhou[1], Yanghan Ou[1], Xiaoyue Fang[1,*], M. Reza Azarpazhooh[2], Haitao Gan[1,*], Zhiwei Ye[1], J. David Spence[2], Xiangyang Xu[3] and Aaron Fenster[2]

[1] School of Computer Science, Hubei University of Technology, Wuhan, China
[2] Robarts Research Institute, Western University, London, Canada
[3] Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

* **Correspondence:** Email: fxy2021@hbut.edu.cn; htgan01@hbut.edu.cn.

Abstract: Carotid total plaque area (TPA) is an important contributing measurement to the evaluation of stroke risk. Deep learning provides an efficient method for ultrasound carotid plaque segmentation and TPA quantification. However, high performance of deep learning requires datasets with many labeled images for training, which is very labor-intensive. Thus, we propose an image reconstruction-based self-supervised learning algorithm (IR-SSL) for carotid plaque segmentation when few labeled images are available. IR-SSL consists of pre-trained and downstream segmentation tasks. The pre-trained task learns region-wise representations with local consistency by reconstructing plaque images from randomly partitioned and disordered images. The pre-trained model is then transferred to the segmentation network as the initial parameters in the downstream task. IR-SSL was implemented with two networks, UNet++ and U-Net, and evaluated on two independent datasets of 510 carotid ultrasound images from 144 subjects at SPARC (London, Canada) and 638 images from 479 subjects at Zhongnan hospital (Wuhan, China). Compared to the baseline networks, IR-SSL improved the segmentation performance when trained on few labeled images (n = 10, 30, 50 and 100 subjects). For 44 SPARC subjects, IR-SSL yielded Dice-similarity-coefficients (DSC) of 80.14–88.84%, and algorithm TPAs were strongly correlated (r = 0.962 − 0.993, $p < 0.001$) with manual results. The models trained on the SPARC images but applied to the Zhongnan dataset without retraining achieved DSCs of 80.61–88.18% and strong correlation with manual segmentation (r = 0.852 − 0.978, p <0.001). These results suggest that IR-SSL could improve deep learning when trained on small labeled datasets, making it useful for monitoring carotid plaque progression/regression in clinical use and trials.

Keywords: carotid plaque; ultrasound image segmentation; self-supervised learning; convolutional neural network

1. Introduction

Stroke is the most common, serious neurological problem globally and the third leading cause of death worldwide, representing staggering mortality, morbidity, and economic cost [1], among which ischemic stroke accounts for the majority. The rupture of atherosclerotic plaques in the carotid artery is one of the most important underlying causes of ischemic stroke [2]. Ultrasound (US) imaging is a widely used standard tool for carotid plaque diagnosis and monitoring because it is non-invasive, low cost, and widely available. Carotid intima-media thickness (IMT) measurement is an early and widely used 2D US biomarker in clinical use [3]. However, IMT is biologically [4], pathologically [5] and genetically [6] distinct from atherosclerosis, and measurement of TPA in the carotid arteries is a method for quantifying atherosclerotic burden, which is a much stronger predictor of stroke risk than IMT, and it is more useful for assessing the effectiveness of therapy [7]. As well, TPA quantification provides a direct method for monitoring the progression and regression of carotid plaques in response to therapy. Obtaining TPA measurement requires the accurate segmentation of each plaque in carotid ultrasound images and summing the plaque areas within the boundaries. This processing is time-consuming. Thus, there is a critical need to develop a 2D US-based automatic carotid plaque segmentation method with good accuracy and reproducibility for monitoring the progression and regression of carotid plaque burden in response to therapy.

Several traditional image processing and machine learning methods have been applied to carotid US image segmentation, such as a snake model [8], a level set method [9], a Bayesian model [10], and a K-means algorithm [11]. However, some of these methods required an initial contour, and were sensitive to the image quality, resulting in the accuracy of the traditional methods to be not satisfactory for clinical use.

Recently, convolutional neural networks (CNNs) have achieved remarkable success in many medical image applications, and some CNNs have been applied to carotid US image segmentation tasks, with the majority applied to automating IMT measurements. Menchon-Lara et al. [12] used a combination of different multi-layer perceptrons to detect the IMT of the common carotid artery from US images . Azzopardi et al. [13] proposed an encoder-decoder convolutional network to segment the lumen-intima boundary (LIB) and the media-adventitia boundary (MAB) by using the envelope and phase congruency data as inputs . Savaş et al. [14] developed a deep learning strategy with multiple hidden layers to segment the IMT from 2D US images. Qian et al. [15] used the continuous max-flow algorithm and a stacked sparse auto-encoder model to extract plaque regions, which were then applied to a random forest classifier for intima-media complex segmentation. Jiang et al. [16] segmented the common and internal carotid arteries from 3D US images using a convolutional neural network (CNN) with adaptive triple loss. In our previous work, we also developed a dynamic convolutional network and a Voxel-FCN network to segment the LIB and MAB from 3D carotid US images [17, 18] and the U-Net++ ensemble algorithm to segment TPA measurement from 2D carotid US images [19].

However, the success of CNNs depends on the use of many labeled images for training, which requires manual delineation of the plaque boundaries. This process is very time-consuming and requires a significant effort of trained individuals due to the low contrast of US images and the variable shapes and appearance of carotid plaques. Training CNN on a small number of labeled images is a particularly challenging task, especially in the application of CNNs for carotid plaque segmentation from 2D US images.

More recently, self-supervised learning (SSL) has been shown to be useful in image classification and segmentation tasks using small sets of labeled images [20, 21]. Zheng et al. [22] proposed a SSL pretext task based on cluster analysis for white blood cell segmentation. Chen et al. [23] used a context-based SSL method to learn semantic features for localization, segmentation, and recognition of brain magnetic resonance images (MRI), abdominal CT images, and fetal ultrasound images. Zhuang et al. [24] proposed a SSL pretext task, named Rubik's Cube, to predict the block orders and rotation angles, and this method was used for brain hemorrhage classification using CT images and brain tumor segmentation of MRI images. Lu et al. [25] used a brain fiber streamline density map to establish a SSL algorithm for brain white matter region segmentation in dMRI images. The above SSL segmentation methods were mainly used for CT and MRI images, which might not be suitable for the carotid US image segmentation task.

The existing deep learning-based methods for ultrasound carotid plaque segmentation were supervised learning, the performance of which relied on a large number of images with well-annotated labels. Although SSL provides a possible way to relieve this problem, most existing SSL methods were more suitable for classification tasks, which could only pre-train the encoder in the segmentation network. In this study, an IR-SSL method is proposed for carotid plaque segmentation from 2D carotid US images and used to study the impact of the use of small, labeled datasets on the accuracy and variability of the segmentation results. IR-SSL can learn the region-wise content information and pre-train parameters in both the decoder and encoder of the downstream segmentation network. The pipeline of IR-SSL consists of a pre-trained stage and a downstream segmentation stage. The pre-trained stage is a different predefined task before the segmentation stage, which aims to learn good feature representations and generate a pre-trained model using the unlabeled carotid US images. Here, we formulate a novel image reconstruction-based task to represent the region-wise content information in the pre-training stage. In the downstream segmentation stage, this pre-trained model is applied to initialize and fine-tune the plaque segmentation network using a small number of labeled carotid US images. Experimental results show that the proposed IR-SSL algorithm could significantly improve the segmentation accuracy of the baseline CNN and achieve better results than the existing segmentation methods when used for training with a small number of labeled images. To the best of our knowledge, this is the first work focusing on SSL for carotid US image segmentation.

2. Materials and method

2.1. Study subjects and image acquisition

The SPARC dataset: In all, 510 US images of carotid plaques were collected from 144 patients who were followed in the Stroke Prevention and Atherosclerosis Research Centre (SPARC, London, Canada) of Western University. The Research Ethics Board of the University approved the use of the anonymized US images (REB number: 113,889). All US images were acquired with an HDI 5000 US system and an L12-5 probe (Philips, Amsterdam, the Netherlands) at a central frequency of 8.5 MHz.

The Zhongnan dataset: A total of 638 images were collected from 497 patients who were followed at Zhongnan Hospital (Wuhan, China). The Zhongnan Hospital Institutional Review Board approved the use of the US images and all patients provided consent. All US images were obtained with an SC2000 US system and a 9L4 probe (Siemens AG, Berlin, Germany) at a central frequency of 6.5 MHz.

All patients included in the experiments had atherosclerotic plaques with risk factors such as hypertension or hyperlipidemia, or a history of vascular events. The carotid plaques were identified in the longitudinal views with a local thickening of the intima greater than 1 mm, as defined in previous studies of TPA [26, 27]. Manual TPAs are generated by two experts via 3D Slicer (www.slicer.org). More details of the two datasets can be found in [19].

2.2. US carotid plaque segmentation via SSL

Training on insufficient labeled data using current deep-learning methods will result in performance degradation. Here, we propose an IR-SSL algorithm to improve the segmentation accuracy of a deep learning algorithm trained on a limited number of labeled images. In this section, we introduce the details of the proposed IR-SSL algorithm. Figure 1 shows the flowchart of the IR-SSL algorithm, including apre-training supervision task and a downstream segmentation task. The pre-training supervision task is formulated by solving an image reconstruction problem using unlabeled carotid ultrasound images. The local representations of the carotid plaque images are learned in this self-supervision stage and then transferred to the segmentation task via the parameter transferring strategy. The downstream segmentation model is then initialized using the pre-trained model and fine-tuned using a small number of labeled images.

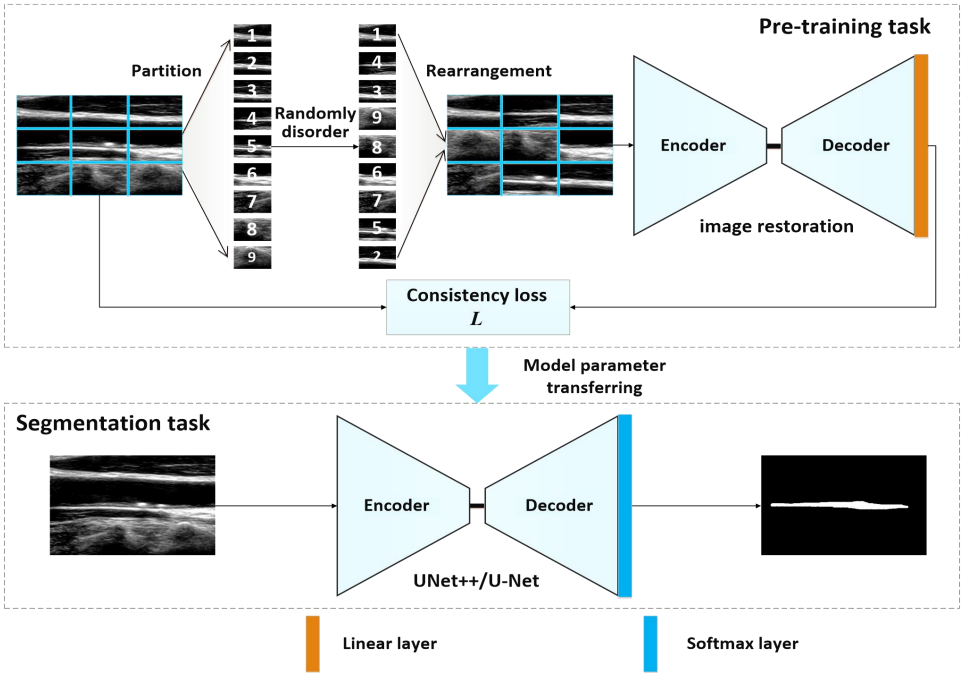

Figure 1. The framework of the proposed IR-SSL algorithm. The upper subgraph is the self-supervised pre-trained task, which is based on the image restoration task, where the image pre-processing step involves three types of operations: partitioning, randomly disordering, and rearranging. The lower subgraph is the segmentation task, in which the pre-trained model is transferred to initialize the segmentation network.

2.2.1. Self-supervision based on image reconstruction

Inspired by the Jigsaw puzzles approach [28], we propose an image reconstruction strategy for self-supervision in the pre-training supervision task, which is simple but useful for the representations of plaque composition and location from unlabeled carotid US images. As shown in the upper subgraph of Figure 1, the self-supervised pre-training task consists of two steps: Generating the rearranged images as the input of the training and reconstructing the original images using an encoder-decoder network.

Given an original dataset $X = [x_1, x_2, \cdots, x_N]$, consisting of N carotid US images without annotated labels, a new dataset with rearranged images is generated by partition, disordering, and rearrangement, such that

$$\widetilde{x} = F(X) \tag{2.1}$$

where $\widetilde{X} = [\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_N}]$ is a set of the rearranged images, $F(\cdot)$ is a transformation function to generate the rearranged images. As shown in Figure 1, the original US image with a plaque is partitioned into N blocks (e.g., $N = 3 \times 3$ with permutations of $(p_1, p_2, ..., p_N)$ and then the permutations of the blocks are disordered (e.g., $(p_1, p_4, ...p_N, ..., p_2)$) and combined to a new rearranged image. Subsequently, the encoder-decoder network learns to approximate the function $G(\cdot)$, which reconstructs the original images X from the rearranged images $\widetilde{X}$, for example

$$Y = G(\widetilde{X}) = G(F(X)) \tag{2.2}$$

2.2.2. Network architecture

The encoder-decoder network can be implemented using different architectures. U-Net and its variants are the most popular networks for medical image segmentation [29]. Thus, in this study, we used the U-Net and UNet++ networks as the basic architectures of our self-supervised segmentation algorithm.

U-Net: the most widely used network for medical image segmentation tasks, consists of a contracting path (encoder) and an expanding path (decoder) [29]. The contracting path has five convolution blocks to extract image features. Each convolution block contains a stack of two convolution operations with the ReLU activity. The expanding path includes four deconvolution blocks. At the end of the network architecture, a 1×1 convolution with a Softmax layer is used to map each 64-component feature vector to a two-channel output.

UNet++: an improved U-Net network, which outperforms variants of the U-Net networks [30]. As shown in Figure 2, UNet++ consists of several U-Nets of varying depths, in which the decoders are densely connected at the same resolution via redesigned skip connections. The loss function is generated by combining losses from these U-Nets with different depths. More details of UNet++ architecture can be found in reference [30].

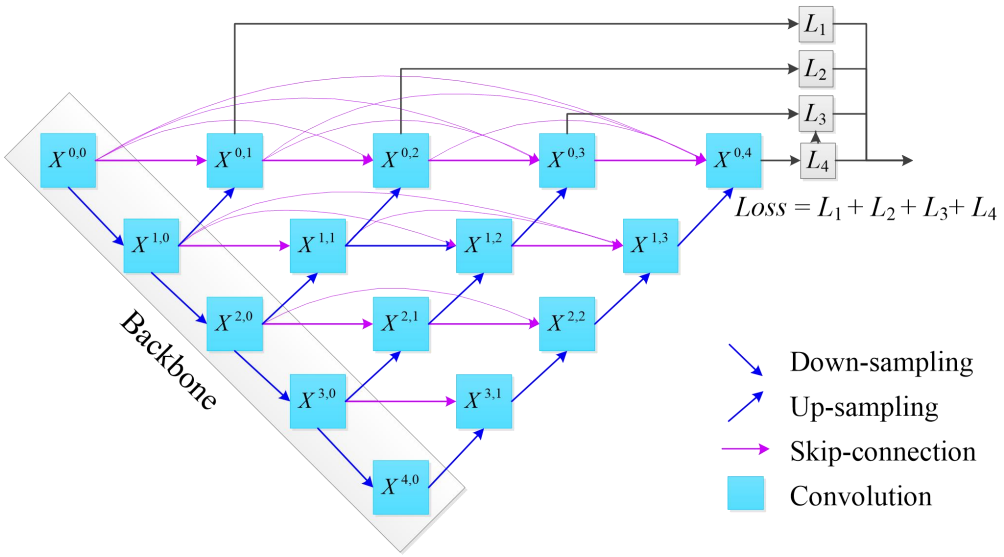

Figure 2. The network architecture of UNet++, consisting of an encoder and decoder, which are connected by a series of nested dense convolution blocks.

The architecture of the encoder-decoder network in the self-supervised network is the same as for the segmentation task but makes use of a linear layer to replace the last Softmax layer. In the self-supervised task, the cosine similarity method is used as the loss function of the encoder-decoder network to minimize the differences between the reconstructed image (Y) and the original image (X), such that

$$\begin{aligned} Loss_{self} &= 1 - sim(X, Y) \\ sim(u, v) &= \frac{u^T v}{\|u\| \cdot \|v\|} \end{aligned} \tag{2.3}$$

where $sim(u, v)$ is the cosine similarity function between two images (u, v).

Cross-entropy was used as the loss function for both the U-Net and UNet++ networks in the segmentation task.

2.3. *Algorithm training and testing*

The training of the IR-SSL algorithm included two stages: self-supervision pre-training and segmentation network fine-tuning. In the self-supervision pre-training stage, the encoder-decoder network pre-training was an unsupervised processing implementation by only using images of the entire dataset without segmentation labels. In the segmentation training stage, the pre-trained model was transferred to the segmentation network via parameter initialization, and the model was then fine-tuned using a small number of labeled images. The SPARC dataset was randomly divided into two subsets as the training and testing datasets. The training dataset included 344 US images with plaques from 100 subjects, and the testing dataset contained 162 images from 44 subjects. To evaluate the performance of IR-SSL after training with limited labeled data sets, we used 10% (R1: n = 10 subjects with 35 plaques), 30% (R2: n = 30 subjects with 103 plaques), 50% (R3: n = 50 subjects

with 172 plaques), and 100% (R4: n = 100 subjects with 344 plaques) of labeled images from the SPARC dataset for training. More importantly, we directly applied the trained models on the SPARC dataset to the Zhongnan dataset without retraining to show the generalizability of the IR-SSL algorithm. The training of the pre-trained task and the baseline U-Net/UNet++ made use of the following parameters: number of epochs = 200, batch size = 8, optimizer = ADAM, learning rate = 0.001, and the training of the segmentation task (fine-tuning) made use of the following parameters: epochs = 50, batch size = 8, optimizer = ADAM, learning rate = 0.0001.

For testing the IR-SSL algorithm, the carotid US images were input to the segmentation network to obtain the plaque boundaries. The TPA measurement was generated by summing the areas within all the segmented plaques from the output of the network for each patient's right and left carotid arteries. Although there is an extra pre-training stage, the self-supervised algorithm does not change the testing flow of deep learning.

Our IR-SSL methods were applied to two popular networks, U-Net and UNet++. We used the standard architecture of U-Net and more details of the architecture can be found in [31]. ResNet152 [32] as the backbone network of UNet++. The input image size for both U-Net and U-Net++ are 96 × 144 pixels. The IR-SSL algorithm and deep learning networks were implemented using Pytorch 1.8.1 and Python 3.9 platforms on an NVIDIA RTX3090 graphics processing unit (GPU).

2.4. Evaluation metrics and statistical analysis

DSC was used to evaluate the overlap area between the algorithm segmentation with the ground truth boundaries of the plaques, as follows:

$$DSC(A, M) = \frac{2 * |A \cap M|}{|A| \cup |M|} 100\% \tag{2.4}$$

The average symmetric surface distance (ASSD) [33] was used to evaluate the distance between the algorithm and manually segmented contours and is given by:

$$ASSD(A, M) = \frac{1}{2}\left(\frac{1}{|\partial R_A|}\sum_{p\in\partial R_A} d(p, \partial R_M) + \frac{1}{|\partial R_M|}\sum_{p\in\partial R_M} d(p, \partial R_A)\right) \tag{2.5}$$

where ∂R_A is the algorithm segmentation surface and $d(p, \partial R_M)$ is the shortest Euclidean distance from a point p to the manually segmented surface ∂R_M; $d(p, \partial R_A)$ are defined in the same manner.

The TPA difference (ΔTPA) was used to quantify the similarity of the algorithm and manually generated TPAs. Continuous variables were expressed as mean ± standard deviation (mean ± SD). The correlation and agreement between manually and algorithm segmented TPA were determined by the Pearson correlation coefficient (r) with a $p < 0.05$ significance level and Bland-Altman plots with 95% limits of agreement (95% LOA) to better visualize the agreement between the two TPA measurements. Graphpad Prism software v7.0 (GraphPad Software, San Diego, CA, USA) was used for the statistical analyses.

3. Results

3.1. Patient characteristics

Table 1 presents the characteristics of 144 patients in the SPARC dataset. The average age is 64 ± 7 years, the average BMI is 29.03 ± 4.99 kg/m, the average systolic blood pressure is 123.08 (13.04) and the average diastolic blood pressure is 72 (8.99).

Table 1. Baseline characteristics of the patients included in the SPARC dataset.

Variable	Value
Age years ± SD	65 ± 7
Sex	Male: 97, Female: 47
Body mass index (kg/m^2)	29.03 ± 4.99
Waist circumference (cm)	101.72 ± 14.09
Smoking current or ex-smokers n, (%)	87 (60.4)
Systolic blood pressure	123.08 (13.04)
Diastolic blood pressure	72 (8.99)
Laboratory data	
Cholesterol (mmol/L)	3.83 (1.22)
Triglyceride (mmol/L)	1.15 (0.60)
High density lipoprotein cholesterol (mmol/L)	1.43 ± 0.41
Low density lipoprotein cholesterol (mmol/L)	1.86 (1.04)
Creatinine (μmol/L)	77.58 (21.22)
TSH (mIU/L)	2.50 ± 1.33
Plasma total homocysteine (μmol/L)	9.74 ± 3.80
Fasting plasma glucose (mmol/L)	5.44 ± 1.63
Hemoglobin A1C	5.98 ± 0.84
Insulin (mU/l)	99.83(70.22)
HOMA-IR	4.31 (5.08)
Medications n, (%)⋆	
Statins	121 (87.1)
Anti-HTN	108 (77.7)
Ezetimibe	76 (54.7)
Fibrate	4 (2.9)
Niacin	16 (11.5)

Data were expressed as n (%) and mean±SD for normally distributed variables and median (interquartile range) for non-normally distributed variables.

HOMA-IR: Homeostatic Model Assessment of Insulin Resistance.

⋆ Missing data: 5 cases.

3.2. Algorithm segmentation results

We performed three sets of experiments to evaluate the proposed IR-SSL algorithm. First, we performed the segmentations using the IR-SSL algorithm with different numbers of partitioned rectangles in the pre-training stage on the SPARC dataset. Second, we showed the details of the results of the IR-SSL algorithm using U-Net and UNet++ architectures on different numbers of labeled training datasets (R1–R4) of the SPARC dataset. Finally, the four trained SPARC models were directly applied to the Zhongnan dataset for testing without retraining to evaluate the generalizability of the IR-SSL algorithm.

3.2.1. Performance with different numbers of partitioned rectangles in IR-SSL

In the supervision stage, the rearranged images were generated from a series of partitioned rectangles, so the number of partitioned rectangles will affect the performance of the IR-SSL algorithm. Here, we compared the performance of the self-supervision algorithm using different numbers of partitioned rectangles (i.e., $N = 2 \times 2$, 3×3, 4×4 and 5×5) on different percentages of labeled training images (i.e., 10, 30, 50 and 100%).

Figure 3 shows the DSC and the SD based on the segmentations generated by the IR-SSL algorithm with the U-Net and UNet++ backbones using different numbers of partitioned rectangles for different percentages of the labeled images used for training. For all cases shown in Figure 3(a),(c), the IR-SSL achieved higher DSCs of (77.22 to 88.84%) and smaller SDs (5.03 to 12.96%) than that of the baseline UNet++ (DSCs = 74.40 to 87.14%, SDs = 6.12 to 13.84%). A similar result is observed in Figure 3(b),(d) for IR-SSL using the U-Net backbone. These results indicate that our self-supervised algorithm can achieve improvements with different numbers of partitioned rectangles in the pre-training stage.

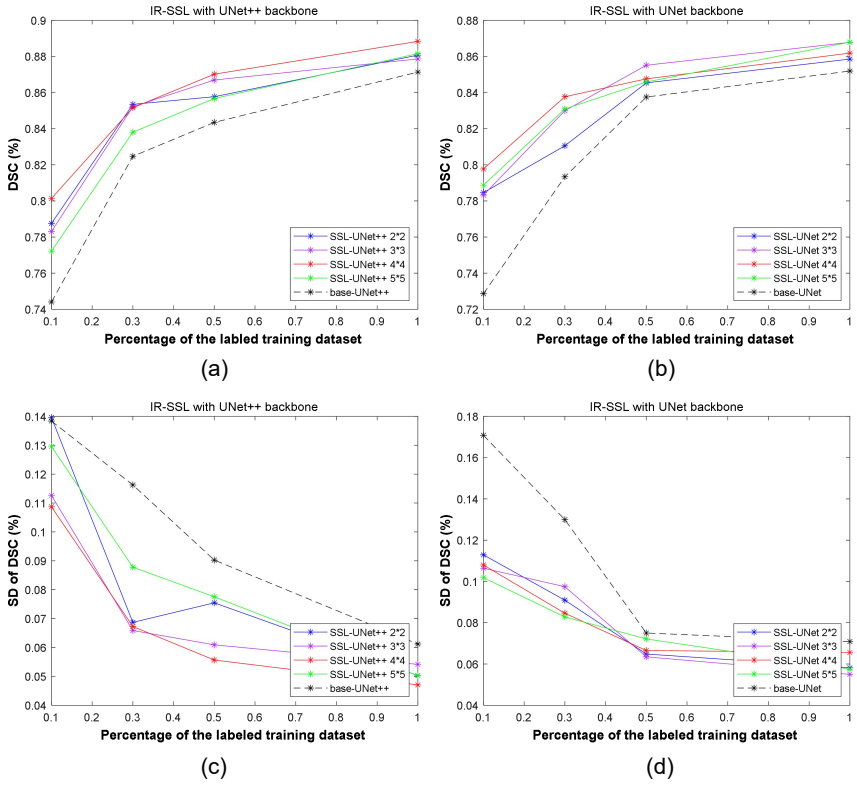

Figure 3. Comparison of the IR-SSL algorithm with different partitioned rectangles to the baseline networks using different percentages of the training images. (a),(b) DSCs of the IR-SSL algorithm with UNet++ and UNet as the backbones on 10, 30, 50 and 100% of the labeled training images. (c),(d) SDs of DSC for the IR-SSL algorithm with UNet++ and UNet backbones trained on 10, 30, 50 and 100% of the labeled training images. The IR-SSL algorithm was implemented with different numbers of partitioned rectangles (blue: $N = 2 \times 2$, purple: $N = 3 \times 3$, red: $N = 4 \times 4$, green: $N = 5 \times 5$) in pre-trained stage.

In most of the cases in Figure 3(a),(c), IR-SSL UNet++ with $N = 4 \times 4$ achieved higher DSC and smaller SD than that of $N = 2 \times 2$, 3×3 and 5×5. In addition, Figure 3(b) shows that IR-SSL UNet with $N = 4 \times 4$ obtained higher DSC for the percentage of training images smaller than 50%, and Figure 3(d) shows that the SDs of IR-SSL UNet were similar for all cases. Thus, we set the number of partitioned rectangles to 16 ($N = 4 \times 4$) in the following experiments.

3.2.2. Performance on datasets with different number of training labels

Figure 4 shows examples of segmented plaque contours using IR-SSL and baseline UNet++ trained on different numbers of labeled images. These results show that the IR-SSL greatly improved the performance of the baseline UNet++ in cases of small numbers of labeled images. Objects 1 and 2 are two successful cases for IR-SSL trained on small numbers of labels, indicating that IR-SSL UNet++ trained on R2 (n = 30 subjects with 103 labeled plaques) could achieve similar to that of the baseline UNet++ trained on R4 (n = 100 subjects with 344 plaques). Objects 3 and 4 are two failure cases on small numbers of training labels. The baseline UNet++ segmentation is rambling on R1 (n = 10 subjects with 35 plaques), but our IR-SSL method could detect the plaque region.

Table 2 shows a comparison of the segmentation performance achieved by the IR-SSL method vs the baseline UNet++ and UNet algorithms trained on various numbers of labeled images (R1: n = 10 subjects with 35 plaques, R2: n = 30 subjects with 103 plaques, R3: n = 50 subjects with 172 plaques, R4: n = 100 subjects with 344 plaques). Compared to the baseline UNet++, the IR-SSL algorithm improved the segmentation DSC by 2.74, 2.69, 2.67 and 1.70%, reduced the ASSD error by 0.14, 0.05, 0.03 and 0.03 mm, and reduced the SDs of ΔTPA measurement by 17.13, 8.49, 6.07 and 3.59 mm^2 for R1, R2, R3 and R4, respectively. Compared to the baseline U-Net, the IR-SSL algorithm improved the segmentation DSC by 5.52, 4.43, 1.01 and 1.00%, reduced the ASSD error by 0.17, 0.05, 0.03 and 0.04 mm, and reduced the SDs of ΔTPA measurement by 9.3, 6.07, 0.06 and 2.95 mm^2 for R1, R2, R3 and R4, respectively. Table 3 also shows the paired t-test between IR-SSL algorithm-generated ΔTPAs and the baseline network-generated ΔTPA on R1, R2, R3 and R4, respectively. As shown in Table 3, paired t-tests showed that the ΔTPA measurements generated by the IR-SSL algorithm were statistically significantly different from that of the baseline UNet++ (R1: $p < 0.011$, R2: $p < 0.172$, R3: $p < 0.0001$, R4: $p < 0.050$) and baseline UNet (R1: $p < 0.019$, R2: $p < 0.0001$, R3: $p < 0.0001$, R4: $p < 0.096$).

Table 2 also presents the comparison of our IR-SSL method to a context-based SSL method reported by Chen et al. [23]. The context-based SSL was implemented with U-Net architecture. Our IR-SSL U-Net achieved better performance than SSL in [23] for R1, R2, R3 and R4 in all metrics. DSC was improved by 0.3–3.0%, ASSD was reduced by 0.08–0.11 mm, and the SD of ΔTPA measurement was reduced by 0.04–3.77 mm^2.

Table 2 and Figure 5 show that IR-SSL UNet++ and IR-SSL UNet yielded TPA measurements trained on all percentages of training datasets that were strongly correlated with the manual segmentations. For all cases, the Bland-Altman plots show good agreement between the IR-SSL UNet++ and manual TPA measurements with a systematic bias of −0.44, −2.46 and 4.29 mm^2 on R1, R2 and R3, respectively. The TPA measurement using IR-SSL UNet++ trained on R2 ($r = 0.979$, bias = −2.46 mm^2, 95%LOA = [−24.6, 19.69] mm^2) achieved similar results as the baseline UNet++ trained on the entire training dataset ($r = 0.983$, bias = −0.64 mm^2 , 95%LOA = [−20.67, 19.39] mm^2). Note that the IR-SSL UNet++ trained on R3 ($r = 0.993$, bias = 4.29 mm^2, 95%LOA =

[−12.82, 24.40] mm^2) was even superior to the baseline UNet++ trained on the entire training dataset.

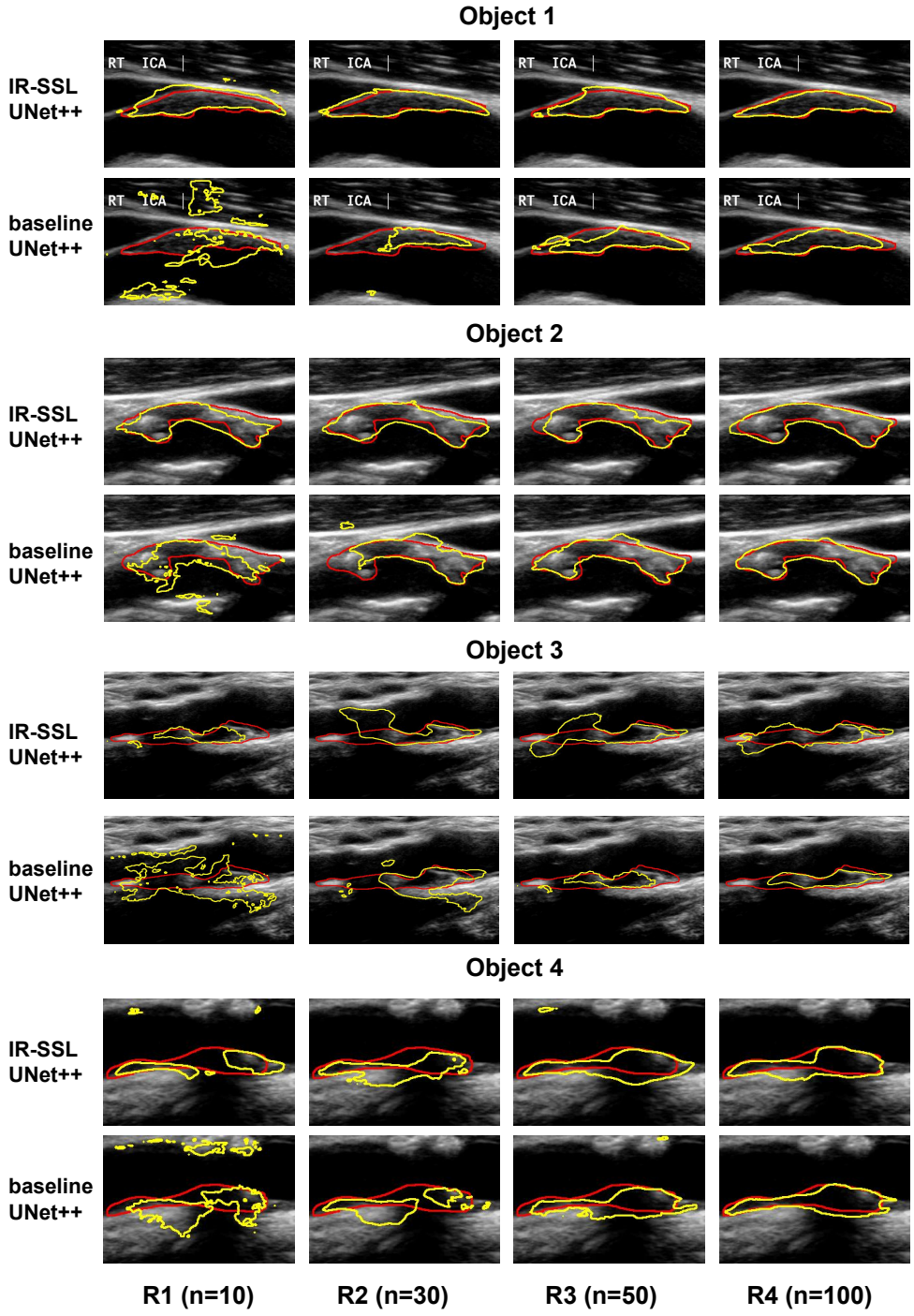

Figure 4. Segmentations of four representative plaques generated by the algorithm (yellow) and manual (red) delineation (SPARC dataset). Each column shows the segmentation generated by the algorithm trained on different percentages of labeled images. For each plaque, the upper row shows the IR-SSL UNet++ segmentations and the lower row shows the baseline UNet++ segmentations.

Table 2. DSC, ASSD, ΔTPA, and Pearson coefficient for plaque segmentation (n = 44 subjects) provided by IR-SSL UNet++, baseline UNet++, IR-SSL U-Net, and baseline U-Net trained on different small labeled datasets (R1: n = 10 subjects, R2: n = 30 subjects, R3: n = 50 subjects, R4: n = 100 subjects).

Metrics (n = 44)	training datasets	DSC (%)	ASSD (mm)	ΔTPA (mm^2)	Pearson *r*
IR-SSL UNet++	R1 (n = 10)	80.14 ± 10.87	0.29 ± 0.21	−0.44 ± 16.37	0.962
	R2 (n = 30)	85.16 ± 6.73	0.22 ± 0.15	−2.46 ± 11.30	0.979
	R3 (n = 50)	87.02 ± 5.57	0.19 ± 0.10	4.29 ± 8.73	0.987
	R4 (n = 100)	88.84 ± 4.74	0.16 ± 0.12	2.31 ± 6.63	0.993
Baseline UNet++	R1 (n = 10)	77.40 ± 14.8	0.43 ± 0.39	10.4 7± 33.50	0.924
	R2 (n = 30)	82.47 ± 11.63	0.27 ± 0.29	−5.43 ± 19.79	0.932
	R3 (n = 50)	84.35 ± 9.03	0.22 ± 0.15	−3.04 ± 14.80	0.965
	R4 (n = 100)	87.14 ± 6.21	0.19 ± 0.20	−0.64 ± 10.22	0.983
IR-SSL U-Net	R1 (n = 10)	78.3 8± 10.11	0.33 ± 0.22	−1.82 ± 17.57	0.947
	R2 (n = 30)	83.77 ± 8.46	0.24 ± 0.14	−0.61 ± 14.17	0.966
	R3 (n = 50)	84.76 ± 6.65	0.21 ± 0.12	3.64 ± 12.66	0.984
	R4 (n=100)	86.19 ± 6.56	0.20 ± 0.15	3.08 ± 9.92	0.973
Baseline UNet	R1 (n = 10)	72.86 ± 17.08	0.50 ± 0.48	10.26 ± 26.87	0.917
	R2 (n = 30)	79.34 ± 13.00	0.29 ± 0.22	−12.47 ± 20.24	0.943
	R3 (n = 50)	83.75 ± 7.50	0.24 ± 0.22	10.99 ± 12.60	0.980
	R4 (n = 100)	85.19 ± 7.09	0.24 ± 0.19	2.35 ± 12.87	0.973
SSL Ref. [23]	R1 (n = 10)	74.40 ± 14.10	0.44 ± 0.36	16.02 ± 21.34	0.944
	R2 (n = 30)	80.80 ± 10.49	0.32 ± 0.29	0.60 ± 14.21	0.965
	R3 (n = 50)	82.25 ± 10.60	0.30 ± 0.31	6.08 ± 13.02	0.977
	R4 (n=100)	85.90 ± 6.17	0.23 ± 0.16	3.21 ± 10.35	0.982

Table 3. The *p* values of paired t-test between ΔTPA generated by IR-SSL method and baseline network on different training datasets (R1: n = 10 subjects, R2: n = 30 subjects, R3: n = 50 subjects, R4: n = 100 subjects).

Methods	Baseline UNet++				Baseline U-Net			
	R1 (n = 10)	R2 (n = 30)	R3 (n = 50)	R4 (n = 100)	R1 (n = 10)	R2 (n = 30)	R3 (n = 50)	R4 (n = 100)
IR-SSL UNet++	0.011	0.172	<0.0001	0.05	–	–	–	–
IR-SSL U-Net	–	–	–	–	0.019	< 0.0001	< 0.0001	0.096

Figure 6 shows the loss and DSC values of training data in each epoch during training, indicating that our IR-SSL method has a lower training loss and a higher DSC than the baseline UNet++ in case of a small number of labeled training data.

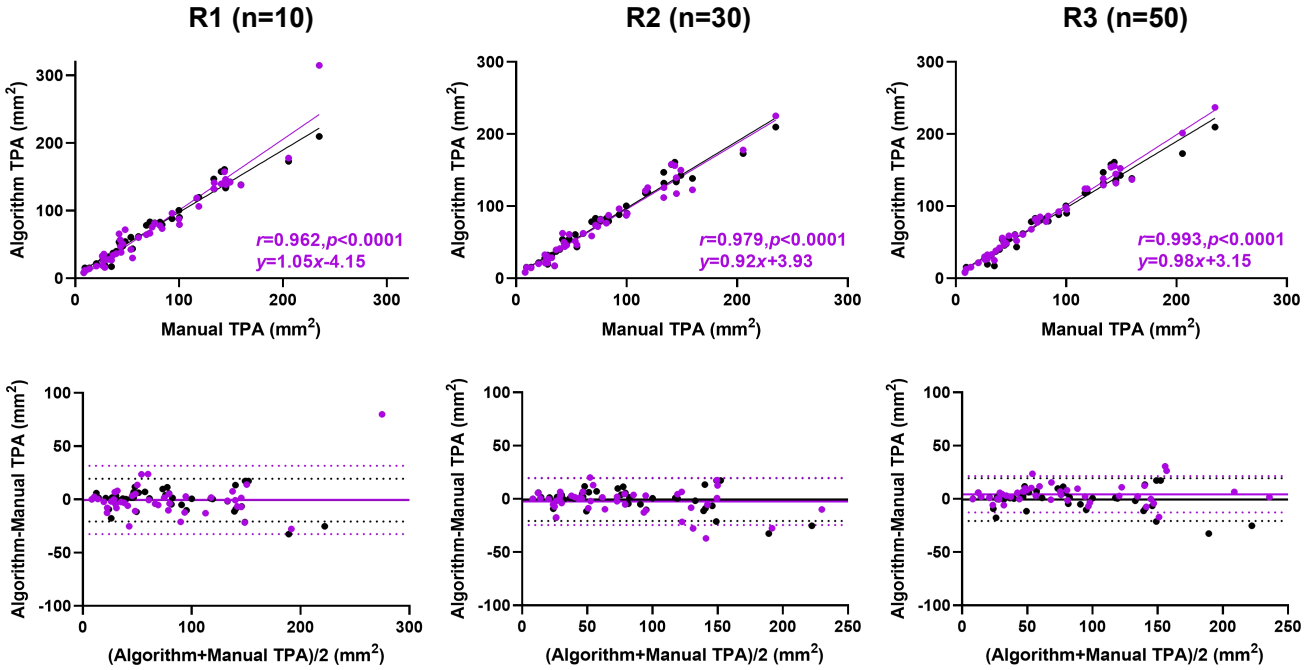

Figure 5. Relationship and agreement of IR-SSL UNet++ and manual TPAs, and baseline UNet++ and manual TPAs. The first row shows the correlations of IR-SSL UNet++ (trained on R1, R2 and R3) and manual TPAs of the test dataset (n = 44, purple), and the correlation of the baseline UNet++ (trained on the entire SPARC training dataset) and manual TPAs (black) with $r = 0.983$, $y = 0.92x + 5.37$, $p < 0.0001$. The second row shows the corresponding Bland-Altman plots of agreement between IR-SSL UNet++ and manual TPAs (R1: bias = −0.44 mm^2, 95%LOA = [−32.52, 31.64] mm^2 ; R2: bias = −2.46 mm^2, 95%LOA = [−24.6, 19.69] mm^2 ; R3: bias = 4.29 mm^2 , 95%LOA = [−12.82, 24.40] mm^2) of the test dataset (n = 44, purple), and baseline UNet++ (percent of training subjects:100%) and manual TPAs of the entire SPARC training dataset (n = 144, black) (bias = −0.64 mm^2, 95%LOA = [−20.67, 19.39] mm^2). The solid lines and the dash lines represent the bias and mean ± 1.96 SD, respectively.

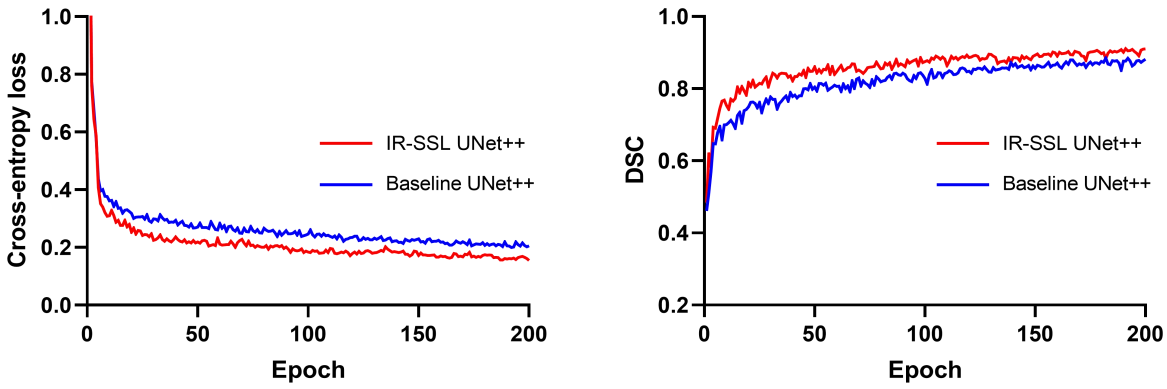

Figure 6. Training loss and DSC curves for R2 (n = 30 subjects).

3.2.3. Generalizability on a different independent dataset

To evaluate the generalizability of the proposed IR-SSL algorithm, we used models trained on the SPARC dataset to predict TPA using a different independent dataset (497 subjects, 638 plaques) that was collected at the Zhongnan hospital (Wuhan, China). Figure 7 shows examples of the IR-SSL UNet++ and manual segmentations of four carotid plaques in the Zhongnan dataset. These results show that the IR-SSL segmentations were similar to the manual segmentations for cases when more than 30 subjects (103 images) in SPARC were used for training.

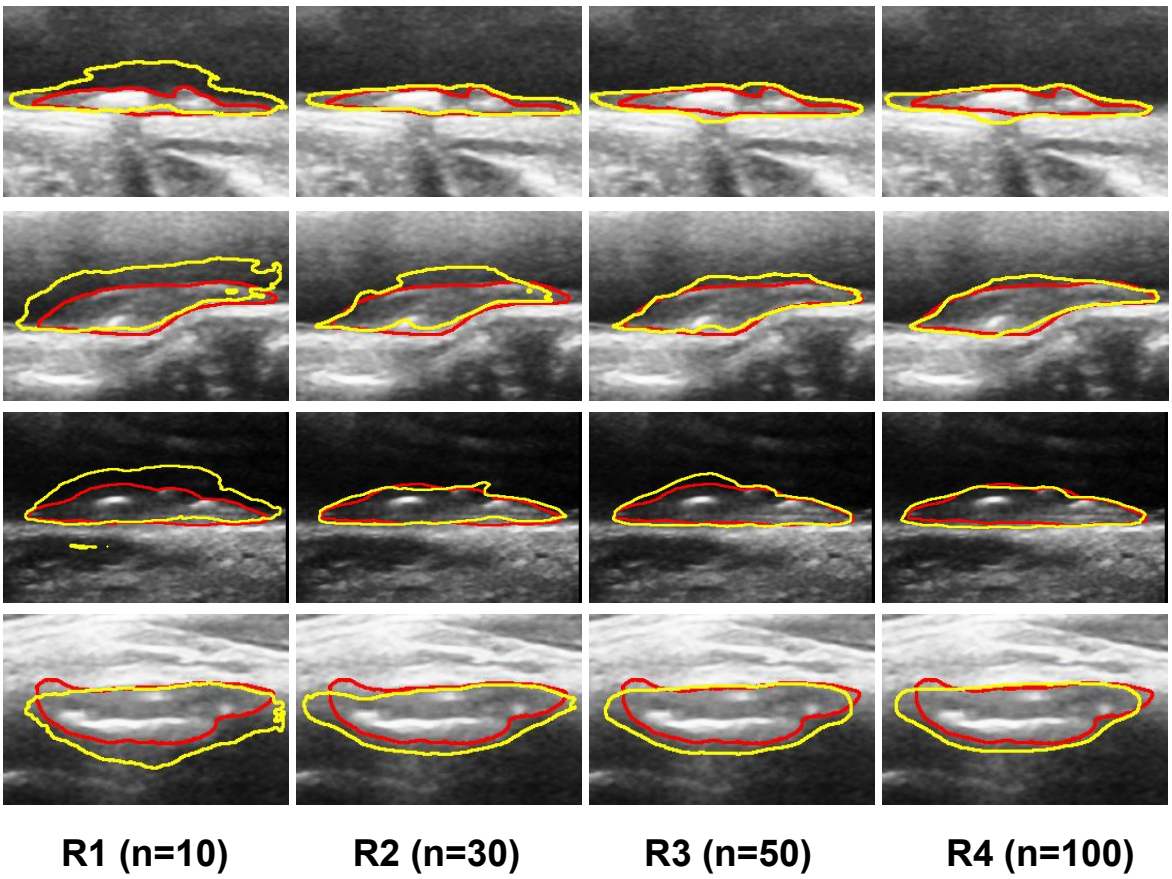

Figure 7. Representative plaque segmentations of three plaques generated by the algorithm (yellow) and manual (red) delineation (Zhonannan dataset). Each row shows the segmentations generated by the algorithm trained on different numbers of labeled images in the SPARC dataset.

Table 4 shows the application of the IR-SSL UNet++ and IR-SSL U-Net models trained on the SPARC dataset (see Sec. 3-B) and used to segment the carotid US images in the Zhongnan dataset. The IR-SSL UNet++ model yielded DSCs of 80.61, 84.91, 85.69 and 88.1%, and ΔTPAs of 9.17, 4.29, 3.78 and 2.73 mm^2 using SPARC models (R1–R4) for prediction. The IR-SSL UNet model also obtained satisfactory results on the Zhongnan dataset. As shown in Figure 8, the IR-SSL UNet++ generated TPA measurements that were highly correlated and agreed with the manual segmentations (R1: $r = 0.852$, $p < 0.000$; R2: $r = 0.936$, $p < 0.0001$; R3: $r = 0.957$, $p < 0.0001$) on the Zhongnan dataset. For

IR-SSL UNet++ trained on R2 and R3 of the SPARC dataset, the differences of algorithm and manual TPAs were relatively small with a bias of −4.30 (95%LOA = [−10.07, 18.68] mm^2) and 3.78mm^2 (95%LOA = [−8.98, 16.55] mm^2).

Table 4. DSC, ΔTPA, Pearson coefficient for plaque segmentation (n = 497 subjects) of the Zhongnnan dataset generated by the IR-SSL UNet++ and IR-SSL U-Net models trained on the SPARC datasets.

Methods	SPARC training datasets	DSC (%)	ΔTPA (mm^2)	Pearson *r*
	R1 (n = 10)	80.61 ± 9.75	9.17 ± 21.02	0.852
IR-SSL	R2 (n = 30)	84.91 ± 6.75	4.29 ± 7.45	0.936
UNet++	R3 (n = 50)	85.69 ± 6.71	3.78 ± 6.52	0.957
	R4 (n = 100)	88.18 ± 4.80	2.73 ± 4.92	0.978
	R1 (n = 10)	81.99 ± 8.76	5.12 ± 12.14	0.874
IR-SSL	R2 (n = 30)	83.30 ± 8.16	5.36 ± 8.47	0.922
U-Net	R3 (n = 50)	86.48 ± 6.49	3.54 ± 7.12	0.937
	R4 (n = 100)	86.90 ± 5.30	2.97 ± 4.59	0.971

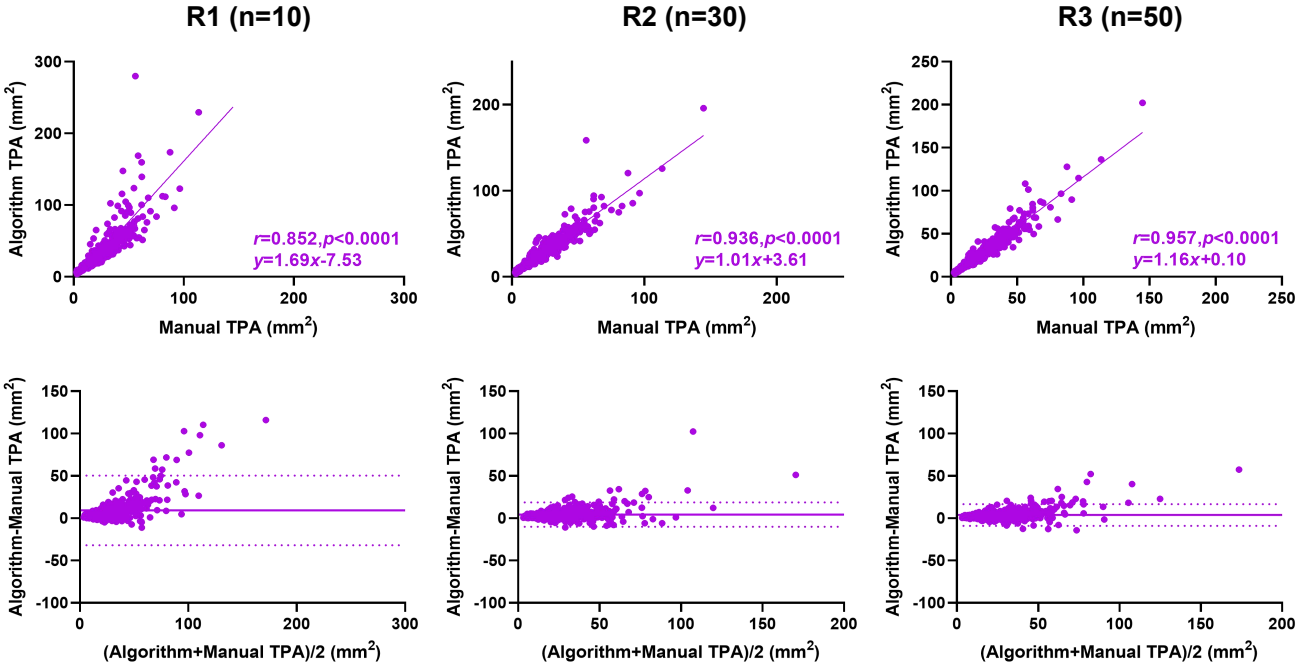

Figure 8. Relationship and agreement of IR-SSL UNet++ and manual TPAs. The first row shows the correlations of IR-SSL UNet++ (trained on R1, R2, R3) and manual TPAs of the test dataset (n = 497). The second row shows the corresponding Bland-Altman plots of agreement between IR-SSL UNet++ and manual TPAs (R1: bias = −0.44 mm^2, 95%LOA = [−32.52, 31.64] mm^2; R2: bias = −2.46 mm^2, 95%LOA = [−24.6, 19.69] mm^2; R3: bias = 4.29mm^2, 95%LOA = [−12.82, 24.40] mm^2) of the test dataset (n = 497). The solid lines and the dash lines represent the bias and mean ± 1.96SD, respectively.

4. Discussion

Carotid TPA is an important US-based measurement, which is used to monitor the progression and regression of carotid plaques, evaluate the effect of medical therapy, and help estimate and stratify the risk for stroke [7, 26]. In this work, we reported on the development of an IR-SSL that can potentially facilitate deep learning-based approaches for automated plaque segmentation and TPA measurement from carotid 2D US images. We implemented our algorithm on two datasets of patients with risk factors (i.e., hypertension or hyperlipidemia, or with a history of vascular events) enrolled at two different centers in different countries, and observed that 1) the pre-trained self-supervision stage could enable CNNs to learn useful image semantics without any labels, which benefits the downstream segmentation task, 2) IR-SSL reduced the number of expert manual annotations that are required for CNN training, and 3) IR-SSL achieved promising segmentation accuracy, reproducibility, and generalizability to an independent dataset of images acquired with a different ultrasound machine in a different institution and country without model retraining.

It is widely recognized that data-driven supervised CNNs require a large number of objects in the training dataset with high-quality expert manual labels. The use of a small dataset for training will result in a degradation of the segmentation performance. To alleviate this degradation, we designed a SSL algorithm based on image reconstruction. Inspired by the existing self-supervised feature learning literature [20, 21], the ideal pre-training task should have a similar goal to the downstream task. Since the location and context semantics is important for the carotid US segmentation task, we, therefore, proposed an image reconstruction self-learning strategy to learn features for these goals in the pre-training stage.

The developed IR-SSL algorithm substantially reduced the need for manual labels for CNN training. Here, we demonstrated the feasibility of the IR-SSL algorithm to train two widely used segmentation networks (UNet++ and U-Net), which were trained using different percentages of labeled images and achieved promising performance of TPA measurements. For both UNet++ and U-Net, the IR-SSL using 103 training labels obtained similar results to the baseline networks using 344 training labels. The IR-SSL using 172 training labels was even superior to the baseline networks using 344 training labels. For example, as shown in Table 1, ΔTPAs provided by IR-SSL UNet++ trained on R2 (n = 30 subjects with 103 images) and R3 (n = 50 subjects with 172 images) were -2.46 ± 11.30 mm^2 and 4.29 ± 8.73 mm^2, and ΔTPA of the baseline UNet++ using the entire labeled training data (n = 100 subjects with 344 images) was -0.64 ± 10.22 mm^2. Furthermore, the IR-SSL algorithm reduced the SD of ΔTPA (Table 2), which is more important in clinical use. The SDs of IR-SSL UNet++ were 6.63–16.37 mm^2; however, the SDs of the baseline UNet++ and U-Net were 10.22–33.50 mm^2. Moreover, the Bland-Altman plots (Figure 5) showed that TPAs of the IR-SSL algorithm trained on a small number of labeled images were significantly correlated and agreed with manual TPAs. We also compared our IR-SSL method to a popular context-based SSL method [23], and IR-SSL achieved better performance than the context-based SSL in all cases of our experiments.

In our previous work [18], we observed obvious performance degradation when applying a trained CNN model to a new and independent dataset (not seen during network training). Our IR-SSL learned plaque location and context semantics from a large number of unlabeled images in the pre-training stage, which might increase the generalizability and reduce the above-described degradation. Here,

we demonstrated the performance of the IR-SSL algorithm trained on the SPARC dataset and tested on the Zhongnan dataset, which are two independent datasets collected from two different countries with different ultrasound systems and patients. Table 4 and Figure 8 showed that the IR-SSL algorithm achieved promising segmentation accuracy and a strong correlation and excellent agreement between algorithm and manual TPA measurements on the Zhongnan dataset using models trained on different percentages of labeled data of the SPARC dataset. This result indicates that it is possible to implement the IR-SSL algorithm in clinical use, which will greatly reduce the work burden of label annotation.

Furthermore, our IR-SSL algorithm involves a parameter of the number of partitioned rectangles in the pre-training stage. We analyzed the effect of the IR-SSL performance by using different numbers of partitioned rectangles from $N = 4$ to $N = 25$. Figure 3 showed that our IR-SSL algorithm could improve the baseline UNet++ and U-Net resulting in higher DSC and lower SD of DSC for different numbers of partitioned rectangles, suggesting that the IR-SSL is insensitive to this parameter.

Here, we also acknowledge some limitations. In this study, we investigated the application of SSL to reduce the number of training labels for deep learning. We found that the performance of our IR-SSL algorithm was greatly decreased when 35 labeled images from 10 subjects were used for training. This phenomenon might be due to the imbalance distribution of plaques when randomly selecting a small number of images for training. We think that a uniform distribution of plaques may be beneficial and improve the performance of deep learning algorithms when using a small number of objects in the training dataset. Although IR-SSL can accept full-size images for segmentation, in our study, ROI images were used as the input of networks, because the image sizes of the plaques ranged from 18 to 480 pixels. Full-size images as the input will cause the network missing small plaques during the down-sampling operations. In the future, we will develop an automatic plaque ROI detection method (e.g., R-CNN), which will address fully this issue. Furthermore, we are planning to extend this work and apply this algorithm to generate 3DUS-based biomarkers of carotid plaque burden in patients at risk for stroke.

5. Conclusions

We propose an IR-SSL that provides a way to utilize a limited labeled training dataset for deep learning and carotid ultrasound plaque segmentation. This algorithm has been compared to the general U-Net++ and UNet, which showed significant improvements when trained on a few labeled images (n =10, 30, 50, and 100 subjects). The proposed IR-SSL method also achieved better performance than the existing SSL method [23]. Furthermore, we directly applied the trained four models on the SPARC dataset (SPARC, London, Canada) to an independent dataset (Zhongnan Hospital, Wuhan, China), and our method achieved satisfactory performance. With high segmentation accuracy, low TPA error, and good generalizability, our method could facilitate the application of deep learning in carotid ultrasound image analysis with limited training labels, and might be suitable for TPA measurement in research and clinical workflow.

Acknowledgments

The authors thank the SPARC research team and Zhongnan hospital research team for providing the 2D US images. This work was supported by the Natural Science Foundation of Hubei Province

under grant No. 2021CFB282, the National Natural Science Foundation of China under grant No. 62201203, the Canadian Institutes of Health Research (CIHR), the High-level Talents Fund of Hubei University of Technology under grant No. GCRC2020016, the Doctoral Scientific Research Foundation of Hubei University of Technology under grant No. BSDQ2020064, the Key R&D Program of Hubei Province under grant No. 2020BAB012, and the National Key R&D Program of China under grant No. 2019YFE0196800.

Conflict of interest

The authors have no conflicts to disclose.

References

1. T. Vos, S. S. Lim, C. Abbafati, K. M. Abbas, M. Abbasi, M. Abbasifard, et al., Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019, *Lancet*, **369** (2020), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9
2. J. D. Spence, Technology insight: Ultrasound measurement of carotid plaque-patient management, genetic research, and therapy evaluation, *Nat. Clin. Pract. Neurol.*, **2** (2006), 611–619. https://doi.org/10.1038/ncpneuro0324
3. M. L. Bots, A. W. Hoes, P. J. Koudstaal, A. Hofman, D. E. Grobbee, Common carotid intima-media thickness and risk of stroke and myocardial infarction: The rotterdam study, *Circulation*, **96** (1997), 1432–143. https://doi.org/10.1161/01.cir.96.5.1432
4. J. D. Spence, Carotid ultrasound phenotypes are biologically distinct, *Arterioscler., Thromb., Vasc. Biol.*, **35** (2015), 1910–1913. https://doi.org/10.1161/ATVBAHA.115.306209
5. A. V. Finn, F. D. Kolodgie, R. Virmani, Correlation between carotid intimal/medial thickness and atherosclerosis: A point of view from pathology, *Arterioscler., Thromb., Vasc. Biol.*, **30** (2010), 177–181. https://doi.org/10.1161/ATVBAHA.108.173609
6. R. L. Pollex, R. A. Hegele, Genetic determinants of carotid ultrasound traits, *Curr. Atheroscler. Rep.*, **8** (2006), 206–215. https://doi.org/10.1007/s11883-006-0075-z
7. E. B. Mathiesen, S. H. Johnsen, T. Wilsgaard, K. H. Bønaa, M. L. Løchen, I. Njølstad, Carotid plaque area and intima-media thickness in prediction of first-ever ischemic stroke: A 10-year follow-up of 6584 men and women: the tromsø study, *Stroke*, **42** (2011), 972–978. https://doi.org/10.1161/strokeaha.110.589754
8. C. Loizou, S. Petroudi, M. Pantziaris, A. Nicolaides, C. Pattichis, An integrated system for the segmentation of atherosclerotic carotid plaque ultrasound video, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **61** (2014), 86–101. https://doi.org/10.1109/tuffc.2014.6689778
9. J. Cheng, H. Li, F. Xiao, A. Fenster, X. Zhang, X. He, et al., Fully automatic plaque segmentation in 3-D carotid ultrasound images, *Ultrasound Med. Biol.*, **39** (2013), 2431–2446. https://doi.org/10.1016/j.ultrasmedbio.2013.07.007
10. F. Destrempes, J. Meunier, M. F. Giroux, G. Soulez, G. Cloutier, Segmentation of plaques in sequences of ultrasonic b-mode images of carotid arteries based on motion

estimation and a bayesian model, *IEEE Trans. Biomed. Eng.*, **58** (2011), 2202–2211. https://doi.org/10.1109/tbme.2011.2127476

11. S. Delsanto, F. Molinari, P. Giustetto, W. Liboni, S. Badalamenti, J. S. Suri, Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images, *IEEE Trans. Instrum. Meas.*, **56** (2007), 1265–1274. https://doi.org/10.1109/TIM.2007.900433
12. R. M. Menchón-Lara, M. C. Bastida-Jumilla, J. Morales-Sánchez, J. L. Sancho-Gómez, Automatic detection of the intima-media thickness in ultrasound images of the common carotid artery using neural networks, *Med. Biol. Eng. Comput.*, **52** (2014), 169–181. https://doi.org/10.1007/s11517-013-1128-4
13. C. Azzopardi, Y. A. Hicks, K. P. Camilleri, Automatic carotid ultrasound segmentation using deep convolutional neural networks and phase congruency maps, in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, (2017), 624–628. https://doi.org/10.1109/ISBI.2017.7950598
14. S. Savaş, N. Topaloğlu, O. Kazcı, P. N. Koşar, Classification of carotid artery intima media thickness ultrasound images with deep learning, *J. Med. Syst.*, **43** (2019), 273. https://doi.org/10.1007/s10916-019-1406-2
15. C. Qian, E. Su, X. Yang, Segmentation of the common carotid intima-media complex in ultrasound images using 2-D continuous max-flow and stacked sparse auto-encoder, *Ultrasound Med. Biol.*, **46** (2020), 3104–3124. https://doi.org/10.1016/j.ultrasmedbio.2020.07.021
16. M. Jiang, Y. Zhao, B. Chiu, Segmentation of common and internal carotid arteries from 3D ultrasound images based on adaptive triple loss, *Med. Phys.*, **48** (2021), 5096–5114. https://doi.org/10.1002/mp.15127
17. R. Zhou, A. Fenster, Y. Xia, J. D. Spence, M. Ding, Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images, *Med. Phys.*, **46** (2019), 3180–3193. https://doi.org/10.1002/mp.13581
18. R. Zhou, F. Guo, M. R. Azarpazhooh, J. D. Spence, E. Ukwatta, M. Ding, A. Fenster, A voxel-based fully convolution network and continuous max-flow for carotid vessel-wall-volume segmentation from 3D ultrasound images, *IEEE Trans. Med. Imaging*, **39** (2020), 2844–2855. https://doi.org/10.1109/tmi.2020.2975231
19. R. Zhou, F. Guo, M. R. Azarpazhooh, S. Hashemi, X. Cheng, J. D. Spence, et al., Deep learning-based measurement of total plaque area in b-mode ultrasound images, *IEEE J Biomed. Health Inform.*, **25** (2021), 2967–2977. https://doi.org/10.1109/jbhi.2021.3060163
20. C. Doersch, A. Zisserman, Multi-task self-supervised visual learning, in *2017 IEEE International Conference on Computer Vision(ICCV)*, (2017), 2070–2079. https://doi.org/10.1109/ICCV.2017.226
21. Z. Ren, Y. J. Lee, Cross-domain self-supervised multi-task feature learning using synthetic imagery, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018). http://dx.doi.org/10.1109/CVPR.2018.00086
22. X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, *Micron*, **107** (2018), 55–71. https://doi.org/10.1016/j.micron.2018.01.010

23. L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara,D. Rueckert, Self-supervised learning for medical image analysis using image context restoration, *Med. Image Anal.*, **58** (2019), 101539. https://doi.org/10.1016/j.media.2019.101539
24. X. Zhuang, Y. Li, Y. Hu, K. Ma, Y. Yang, Y. Zheng, Self-supervised feature learning for 3D medical images by playing a rubik's cube, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2019), 420–428. https://doi.org/10.1007/978-3-030-32251-9_46
25. Q. Lu, Y. Li, C. Ye, White matter tract segmentation with self-supervised learning, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2020), 270–279. https://doi.org/10.1007/978-3-030-59728-3_27
26. J. D. Spence, M. Eliasziw, M. Dicicco, D. G. Hackam, R. Galil, T. Lohmann, Carotid plaque area: A tool for targeting and evaluating vascular preventive therapy, *Stroke*, **33** (2002), 2916–2922. https://doi.org/10.1161/01.str.0000042207.16156.b9
27. S. H. Johnsen, E. B. Mathiesen, O. Joakimsen, E. Stensland, T. Wilsgaard, M.L. Løchen, et al., Carotid atherosclerosis is a stronger predictor of myocardial infarction in women than in men: A 6-year follow-up study of 6226 persons: the tromsø study, *Stroke*, **38** (2007), 2873–2880. https://doi.org/10.1161/strokeaha.107.487264
28. M. Noroozi, P. Favaro, Unsupervised learning of visual representations by solving jigsaw puzzles, in *European Conference on Computer Vision*, (2016), 69–84. https://doi.org/10.1007/978-3-319-46466-4_5
29. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
30. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, *IEEE Trans. Med. Imaging*, **39** (2020), 1856–1867. https://doi.org/10.1109/tmi.2019.2959609
31. R. Zhou, M. R. Azarpazhooh, J. D. Spence, S. Hashemi, W. Ma, X. Cheng, et al., Deep learning-based carotid plaque segmentation from b-mode ultrasound images, *Ultrasound in Medicine & Biology*, **47** (2021), 2723–2733. https://doi.org/10.1016/j.ultrasmedbio.2021.05.023
32. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–778. https://doi.org/10.1109/CVPR.2016.90
33. T. Heimann, B. V. Ginneken, M. A Styner, Y. Arzhaeva, V. Aurich, C. Bauer, et al., Comparison and evaluation of methods for liver segmentation from CT datasets, *IEEE transactions on medical imaging*, **28** (2009), 1251–1265. https://doi.org/10.1109/TMI.2009.2013851

AIMS Press