

OF-MSRN: Optical Flow-Auxiliary Multi-Task Regression Network for Direct Quantitative Measurement, Segmentation and Motion Estimation

Chengqian Zhao,¹ Cheng Feng,² Dengwang Li,^{1*} Shuo Li^{3,4}

¹Shandong Key Laboratory of Medical Physics and Image Processing & Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine, School of Physics and Electronics, Shandong Normal University, Jinan, Shandong 250358, China

²Department of Ultrasound, The Second Affiliated Hospital, Southern University of Science and Technology, Shenzhen Third People’s Hospital, Shenzhen, Guangdong, China

³Department of Medical Imaging, Western University, London, ON, Canada, ⁴Digital Image Group, London, ON, Canada
chengqian.zhao@139.com, lidengwang@sdnu.edu.cn

Abstract

Comprehensively analyzing the carotid artery is critically significant to diagnosing and treating cardiovascular diseases. The object of this work is to simultaneously achieve direct quantitative measurement and automated segmentation of the lumen diameter and intima-media thickness as well as the motion estimation of the carotid wall. No work has simultaneously achieved the comprehensive analysis of carotid artery due to three intractable challenges: 1) Tiny intima-media is more challenging to measure and segment; 2) Artifact generated by radial motion restrict the accuracy of measurement and segmentation; 3) Occlusions on diseased carotid walls generate dynamic complexity and indeterminacy. In this paper, we propose a novel optical flow-auxiliary multi-task regression network named OF-MSRN to overcome these challenges. We concatenate multi-scale features to a regression network to simultaneously achieve measurement and segmentation, which makes full use of the potential correlation between the two tasks. More importantly, we creatively explore an optical flow auxiliary module to take advantage of the co-promotion of segmentation and motion estimation to overcome the restrictions of the radial motion. Besides, we evaluate consistency between forward and backward optical flow to improve the accuracy of motion estimation of the diseased carotid wall. Extensive experiments on US sequences of 101 patients demonstrate the superior performance of OF-MSRN on the comprehensive analysis of the carotid artery by utilizing the dual optimization of the optical flow auxiliary module.

Introduction

Comprehensive analysis of carotid artery based on direct quantitative measurement, automated segmentation, and motion estimation is significant to diagnosing and treating cardiovascular diseases. A series of clinical researches (Alan

*Corresponding author. This work was funded by the National Natural Science Foundation of China (61971271), the Taishan Scholars Project of Shandong Province (Tsqn20161023) and the Primary Research and Development Plan of Shandong Province (No. 2018GGX101018).

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

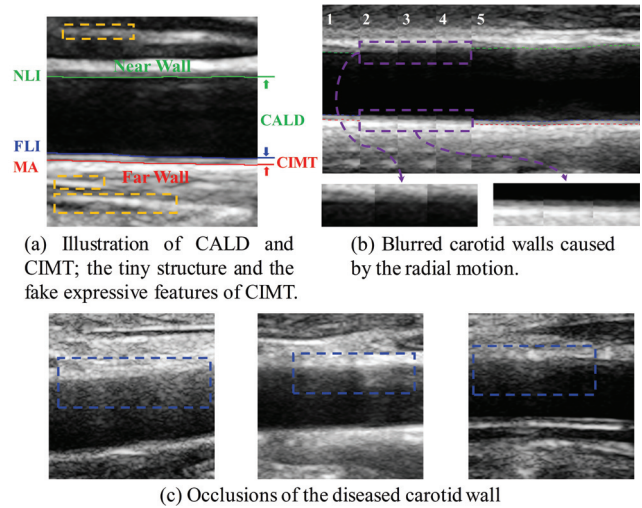


Figure 1: Challenges of simultaneously achieving direct quantitative measurement, segmentation and motion estimation of the carotid artery. The tiny structure of CIMT, radial motion, and diseased carotid wall cause the loss of expression features and discriminative information.

et al. 2003; Bots et al. 2005; Naqvi and Lee 2014) have verified that carotid artery lumen diameter (CALD) and intima-media thickness (CIMT) are risk indicators for many cardiovascular diseases. The CALD is the distance between lumen-intima interfaces at the near wall (NLI) and far wall (FLI), and the CIMT is the distance between FLI and media-adventitia interface at the far wall (MA). Bots et al. (Bots, Hofman, and Grobbee 1997) have indicated that CIMT is related to changes in local shear stress and tensile stress and maybe a direct function of CALD. Simultaneously quantifying CALD and CIMT is great significance. Besides, increased stiffness of the carotid artery is significantly correlated to the incidence of cardiovascular diseases. Estimating the periodical motion of the carotid wall is the basis for analyzing the stiffness. In arterial wall dynamics, CALD and CIMT are crucial factors in evaluating carotid artery elas-

ticity, and estimating radial motion is the basis for analysis arterial stiffness (Baltgaile 2012). Therefore, comprehensive analysis is significant to provide reliable and objective auxiliary diagnoses of cardiovascular diseases and relieve clinicians from laborious workloads.

No work has achieved comprehensive analysis of carotid artery due to the following exceedingly intractable challenges: (1) CIMT has tiny anatomical structure (Figure 1 (a)) that results in many fake samples (yellow boxes in Figure 1 (a)). It is difficult to extract the correct structure features of the CIMT, which seriously reduces the accuracy of measurement and segmentation. (2) The radial movement of the carotid artery results in blurred boundaries. Artifact (purple boxes in Figure 1 (b)) generates on NLI and FLI during the diastolic motion. It causes large errors for the segmentation. (3) Diseased tissue that covering the carotid wall severely occludes NLI and FLI (blue boxes in Figure 1 (c)). Boundaries of the diseased carotid wall are extremely difficult to find. The US images of these patients lose much discriminative information so that measurement, segmentation and motion estimation are suffered from seriously affect.

We propose a novel optical flow-auxiliary multi-task regression network named OF-MSRN (Figure 2) to achieve direct quantitative measurement, segmentation and motion estimation of carotid artery simultaneously. The proposed OF-MSRN consists of direct quantitative measurement and segmentation branch (DQMS) and unsupervised motion estimation branch (ME). Two branches share a Siamese style multi-scale feature encoder (MFE) to extract the tiny structural feature of CIMT. Especially, we design an OF auxiliary module (OFA) shared by DQMS and ME to simultaneously optimize segmentation and motion estimation. OF-MSRN obtains not only pathological correlations of CALD, CIMT and periodical motion of carotid wall but also global structural information of carotid artery. Therefore, OF-MSRN has great clinical application prospects due to its comprehensive analysis and higher consistency. Its main contributions are as follows:

(1) For the first time, direct measurement, segmentation, and motion estimation are achieved simultaneously. OF-MSRN not only fully take account of the pathological correlation between multiple objects (CALD, CIMT, and carotid wall) but also make full use of co-promotion of multiple tasks. Thus, OF-MSRN provides a reliable, comprehensive and objective auxiliary diagnosis for cardiovascular diseases.

(2) Our newly proposed OFA successfully achieves co-promotion between segmentation and motion estimation by dual optimization. Therefore, the artifact interference caused by radial motion is effectively handled.

(3) The bidirectional OF is adopted to resolve inconsistencies caused by diseased tissue. Furthermore, an OF pyramid is applied to train unsupervised learning. Thus, OF-MSRN is able to effectively handle the disturbance of occlusions for motion estimation.

Related Work

Active contours, dynamic programming, and integrated approaches for segmenting the carotid wall and tracing

the boundaries are overwhelmingly not good in all aspects (Naik, Gamad, and Bansod 2013). And existing deep learning-based works on carotid artery only focus either on CALD or CIMT or only on measurement or segmentation. Biswas et al. (Biswas et al. 2019) first attempt to segment and measure the CALD utilizing the fully connected network (FCN) and showed considerable improvement in accuracy compared to the traditional automated algorithms. Shin et al. (Shin et al. 2016) propose a convolutional neural network (CNN)-based system that performs reliably in frame selection, the region of interest localization, and CIMT measurements. Biswas et al. (Biswas et al. 2018) propose the same intelligence-based CIMT measurement utilizing a four-phases deep learning network and achieve accurate segmentation of FLI and MA. These works effectively calculate the segmentation results to obtain the values of measurement, and they demonstrated the close correlation between segmentation and measurement. Despite the above works have made great contributions to the segmentation and measurement of the carotid artery, all of them regard measurement value as post-processing of neural network rather than as part of it. Their almost entire dependence on segmentation which still is a huge challenge in medical image analysis limits the accuracy of measurement and hinders them from efficient clinical application.

Although many deep learning-based methods have great performance in motion estimation, none of them pay attention to the study of the carotid artery. Golemati et al. (Golemati et al. 2012) have confirmed that the OF is an effective technique for estimating the periodical motion of the carotid artery. A series of works (Dosovitskiy et al. 2015; Ilg et al. 2017; Sun et al. 2018) have achieved motion estimation by deep learning-based OF prediction, and Fischer et al. (Dosovitskiy et al. 2015) creatively proposed the color coding for visualizing motion fields. Recently, many works (Tsai, Yang, and Black 2016; Cheng et al. 2017; Qin et al. 2018) have applied OF prediction-based motion estimation to medical image segmentation and indicated the co-promotion between segmentation and motion estimation. Therefore, OF prediction is not only applicable to the motion estimation of the carotid artery but also can effectively handle the artifact interference during the radial motion and improve the segmentation accuracy.

OF-Auxiliary Multi-Task Regression Network

OF-MSRN simultaneously achieves direct quantitative measurement and segmentation of CALD and CIMT as well as motion estimation of carotid periodic motion. OF-MSRN adopts the Siamese style MFE to effectively extract the tiny structural feature of CIMT for all tasks. The infrastructure of the OF-MSRN is the **multi-task DQMS** (subsection 1), in which the **multi-task regression network** (MRN, Figure 2 (a)) regards the measurement and segmentation tasks as a regression prediction at the pixel level. The MRN predicts the rough segmentation results in the term of coordinates. Another necessary part is the **unsupervised ME** (subsection 2), in which the **BOF** (Figure 2 (b)) adopts a laterally concatenated optical flow (OF) decoder to predict a bi-directional OF pyramid for the motion estimation. The

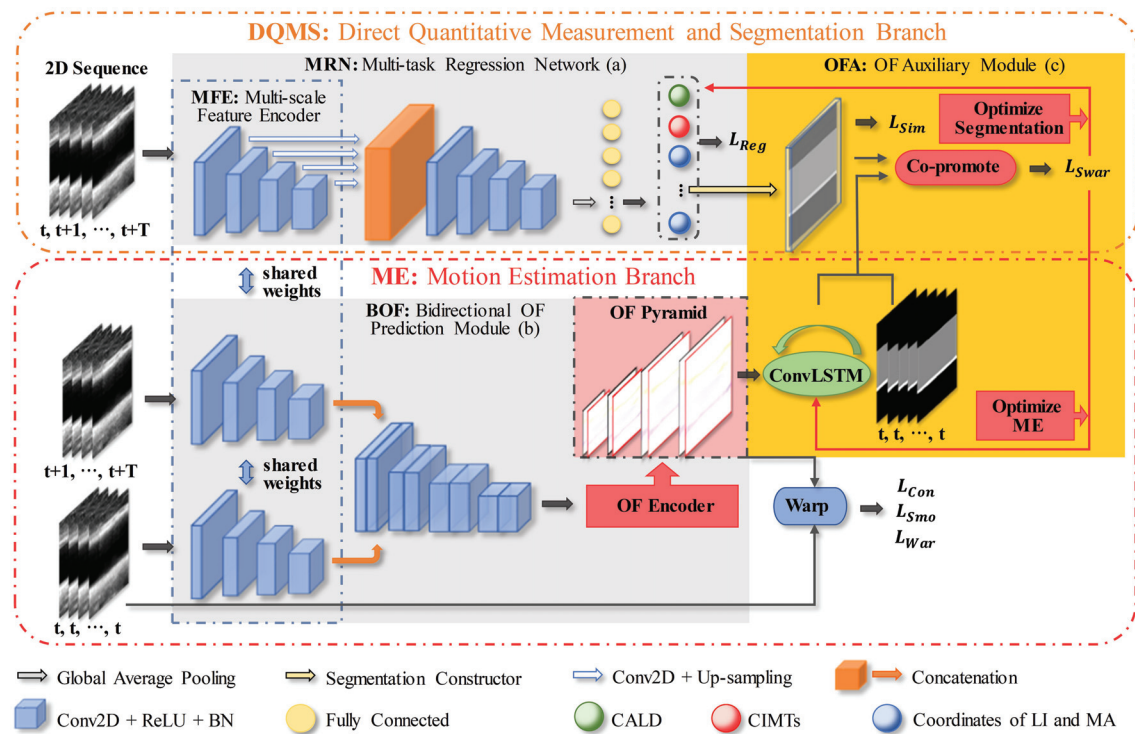


Figure 2: Overview of OF-MSRN. The innovative DQMS and ME share the Siamese style MFE and the OFA. (a) The multi-task regression network (MRN) for direct quantitative measurement and segmentation. (b) The bidirectional OF prediction module (BOF) for motion estimation. (c) OFA to optimizes segmentation and motion estimation simultaneously. Benefit from the newly proposed OFA, OF-MSRN takes advantage of co-promotion between segmentation and motion estimation and achieves an accurate comprehensive analysis of carotid artery.

bi-directional mechanism not only reduces the interference of the artifact to measurement and segmentation but also ensures the accuracy of OF prediction at the occlusions of the diseased carotid wall. Especially, we creatively propose the **OFA** (Figure 2 (c), subsection 3) to optimize segmentation and motion estimation simultaneously. Firstly, OFA uses a constructor to generate the segmentation image from the prediction of MRN. At the same time, OFA propagates OF information along the spatiotemporal dimension by ConvLSTM to capture the motion fields. Finally, OFA wraps the constructed image utilizing predicted motion fields. Joint optimization of segmentation and motion estimation is to improve the similarity between the warped images and the manually segmented images. Assisted by OFA, the segmentation is optimized twice, and the motion estimation is optimized for the second time. OF-MSRN effectively extracts the structural features of CIMT as well as overcomes the influence of artifact and diseased carotid wall. It obtains not only pathological correlations of CALD, CIMT and periodical motion of carotid wall but also global structural information of the carotid artery.

DQMS: Multi-task Regression

Simultaneous implementation of measurement and segmentation is summarized as a multi-output regression. Given

a dataset $S = \{(x_i, y_i) | i = 1, \dots, N\}$, our goal is training a multi-output regression model to learn the mapping $f : x \in R^{h \times w} \rightarrow y \in R^d$, where x_i and y_i denote the US image and the corresponding multiple ground truth respectively, and N is the number of training samples. The DQMS regards NLI, FLI, and MA as a series of discrete points described by coordinates: $C_i = \{(h_i, w_i) | i = 1, \dots, n\}$, where C_i are the point sets of NLI, FLI or MA, h_i and w_i are the i^{th} point's coordinate along with height and width respectively, and n is the number of discrete points of each boundary. To smoothly approximate boundary with complicated and diverse shapes, and keep lower computation, we empirically make n equal 64. We use the spline method to obtain these discrete boundary points from left to right at equal intervals along each boundary. The w_i of all points satisfy the function: $w_i = ai - a + i - 1$, where a is the interval size between two adjacent points. Thus, we regard the w_i as a known parameter and only predict the h_i . Our boundary representation method has few assumptions and flexible structure. Therefore, our regression method provides a more reliable and accurate boundary representation when handling carotid US images with wide shape variations caused by different subjects and pathologies. Simultaneous implementation takes advantage of the underlying co-promotion between measurement and segmentation.

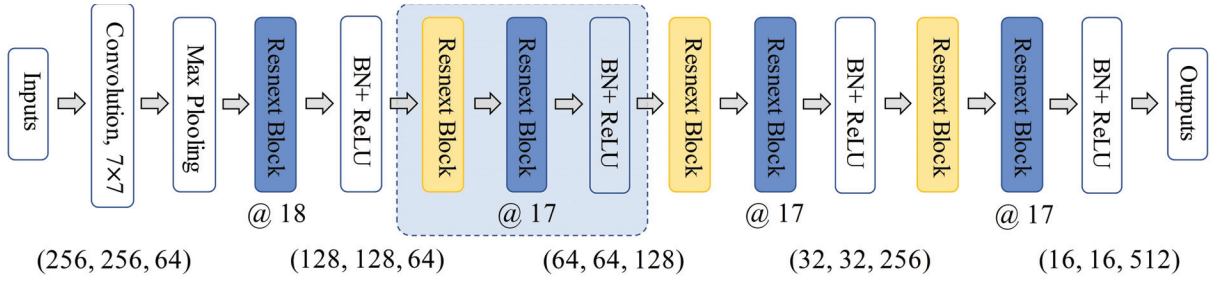


Figure 3: Overview of the MFE. The yellow Resnext blocks does down-sample and the blue doesn't. '@' denotes the number of the blocks. The blue dotted box denotes a feature coding module, which contains 18 Resnext blocks, a batch normalization layer and a ReLU activation layer. MFE contains four feature coding modules. The position of brackets at the bottom, in addition to the first, denotes the position of feature map extraction. The numbers in brackets denote height, width, and number of channels of the feature maps respectively.

MRN simultaneously predicts the vales of CALD, CIMT and their boundary coordinates. Firstly, MFE (Figure 3) obtains more detailed information relative to the top-level features. It adopts a convolution layer to extract the low-level features, and then embed a max-pooling layer to retain main features and increase the receptive field of the convolution kernel. The feature coding module cascades 18 Resnext (Xie et al. 2017) blocks and takes advantage of the batch normalization and the ReLU to improve the adaptability and prevent overfitting. We design a downsampling Resnext block and standard Resnext blocks differ in the first convolution layer and the residual branch. Benefit from the stacked structure of Resnext, the MFE adopts the 'split-transform-merge' strategy of Inception (Szegedy et al. 2017) and the 'stack' strategy of VGG (Simonyan and Zisserman 2014). Therefore, the MFE is able to improve the accuracy without increasing the complexity of parameters. Then, MRN merges the detailed information of multi-scale feature maps and extract the most effective information utilizing four cascaded down-sampling Resnext blocks. Finally, MRN adopts the linear regression to learn an expressive feature embedding and a reliable regressor. Based on the top-level feature $h(x_i)$, outputs of the linear regression are calculated as: $f(x_i) = w_r h(x_i) + b_r$, where w_r and b_r are the weights matrix and bias of the linear regression respectively. Like most regression networks, we optimize our MRN by optimizing the mean square error (MSE) utilizing the ground truth. Benefit from multi-scale information merging and dual selection, the MRN reduces the variability of inter-subjects as well as simplify the relationship between the latent feature space and target output manifold.

ME: Unsupervised Motion Estimation

ME provides accurate OF prediction for carotid motion estimation. We propose a novel BOF (Figure 2 (b)) to unsupervised estimate the periodical motion of carotid artery. In carotid US sequences, BOF predicts a sequence of consec-

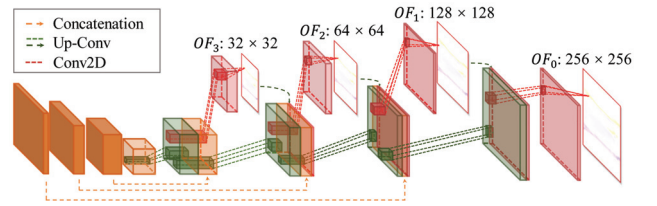


Figure 4: Illustration of the OF encoder. A four-layer pyramid is predicted, above which the numbers denote the OF size.

utive OF representations between target and source frames. The BOF shares the Siamese style MFE with the MRN to encode multi-scale feature maps, and concatenates these features in channel to predict OF_{fwd} and OF_{bwd} . The bidirectional OF share the OF encoder (Figure 2 (b)) whose main ingredient is up-convolution layers consisting of up-sampling and convolution. For brevity, we mainly illustrate the cases from target to source frames in the following, which one can easily generalize to the reversed cases. As illustrate in Figure 4, BOF applies the up-convolution to feature maps and concatenate to the feature maps at the next scale. Concatenated features are convolved to predict the coarser OF. Then the up-sampled coarser OF joint in the next concatenation. These steps are repeated 4 times to obtain a four-layer bidirectional OF pyramids (Figure 2 (b)). BOF preserves both the high-level information passed from coarser OF and detailed structure information in lower feature maps. Each step increases the resolution twice. Therefore, BOF captures the micron-scale axial OF and gets more accurate motion vectors.

To optimize the unsupervised BOF, we utilize the image construction objectively as supervision with the assumption of photometric consistency. In a temporal continuous frame $I_i (i = t, \dots, t + T)$, the target frame I_t is specified as reference image, and the others are source frames $I_{t+k} (k = 1, \dots, T)$. Then, we scale them to build the target image pyramid I_t^s and source image pyramid I_{t+k}^s . The s denotes the s^{th} layer, here $s = 4, 3, 2, 1$. We apply *differen-*

table inverse warping (Jaderberg et al. 2015) to the I_{t+k}^s and evaluate the similarity between the warped image pyramid and the I_t^s . We use a combination of an $L1$ and multi-scale structural similarity (SSIM) (Wang et al. 2004) for the OF warping loss L_{War} which maintains the balance between appropriate assessment of perceptual similarity and modest resilience for outliers, and is differentiable as follows:

$$L_{War}^{s-fwd} = \frac{1}{HW} \sum_{hw} \alpha \frac{1-SSIM(I_t^s, \hat{I}_t^{s-fwd})}{2} + (1-a) \|I_t^s - \hat{I}_t^{s-fwd}\|_1, \quad (1)$$

$$L_{War} = \sum_{s=1}^4 (L_{War}^{s-fwd} + L_{War}^{s-bwd}). \quad (2)$$

where L_{War}^{s-fwd} and L_{War}^{s-bwd} are loss functions of source-to-target and target-to-source warping, \hat{I}_t^{s-fwd} denote the fake target image, $SSIM$ denotes the structural similarity index and α is taken to be 0.85 by cross-validation. In this way, we optimize each layer of the OF pyramid and get the most detailed original scale OF_{fwd} . It is of great significance for optimizing segmentation and motion estimation.

To ensure local smoothness, we penalize the gradients by an edge-aware OF smoothness loss L_{Smo} . We calculate the smoothness loss L_{Smo}^{s-fwd} and L_{Smo}^{s-bwd} respectively at each scale, and then sum them up:

$$L_{Smo}^{s-fwd} = \frac{1}{HW} \sum_{hw} |\nabla OF_{fwd}(p_{hw})| \cdot (e^{-|\nabla I_t(p_{hw})|}), \quad (3)$$

where $|\cdot|$ denotes elementwise absolute value, ∇ is the vector differential operator including ∂_h and ∂_w , p_{hw} denotes coordinates of pixels in OF and image, and T denotes the transpose of image gradient weighting. Be the same as Eq. (2), the L_{Smo} is the sum of smoothness loss of the bidirectional OF. Therefore, we can effectively filter out preserve sharp details and erroneous predictions.

Though ME performs well in most sequences benefiting from robust image similarity assessment (Eq. (1)) and smoothness loss (Eq. (3)), OF fields of the diseased carotid wall still cannot be satisfactorily predicted in practice. To further mitigate these effects, we apply an adaptive forward-backward consistency loss (Yin and Shi 2018) L_{Con} that is optimized to enforce the geometric consistency:

$$L_{Con}^{s-fwd} = \frac{1}{HW} \sum_{hw} \Phi(p_{hw}) \cdot \|OF_{fwd}(p_{hw})\|_1, \quad (4)$$

where $\Delta OF_{fwd}(p_{hw})$ is the difference of OF_{fwd} calculated by forward-backward consistency evaluation at the pixel p_{hw} in I_t , $[\cdot]$ is the Iverson bracket and $\Phi(p_{hw})$ denotes the consistency condition of

$$\|\Delta OF_{fwd}(p_{hw})\| < \max\{\alpha, \beta\|OF_{fwd}(p_{hw})\|_2\}, \quad (5)$$

and we empirically set (α, β) to be (1.0, 0.05). Be the same as Eq. 2, the L_{Con} is the sum of consistency loss of the bidirectional OF. Pixels where the OF_{fwd} and OF_{bwd} contradict seriously are considered as possible outliers. Benefit from the consistency evaluation, we can effectively handle the disturbance of the occlusion and obtain accurate motion fields

of the diseased carotid wall. The summarize loss of the BOF can be defined as: $L_{OF} = \lambda_w L_{War} + \lambda_s L_{Smo} + \lambda_c L_{Con}$, where λ_s (set as 1) denote respective loss weight. Benefit from the optimization of structural similarity, smoothness, and consistency, the erroneous prediction is filtered pixel-by-pixel. Therefore, we can extract a precise dense OF_{fwd} for representing motion field and optimizing segmentation.

Optical Flow-Auxiliary

OFA is to achieving co-promotion between segmentation and motion estimation (Figure 2 (c)). OFA needs two necessary conditions: segmentation images \tilde{I} and motion fields. We design an image constructor to obtain \tilde{I} . The constructor regards the image to be constructed as an array that masked by a zero array. OFA reads the discrete h coordinates from MRN and calculates the corresponding w coordinates to determine the NLI, YLI, and MA, and then generates the \tilde{I} . OFA applies an evaluation such as Eq. (1) to optimize similarity loss L_{Sim} between the \tilde{I} and manually segmented labels L_i ($i = t, \dots, t + T$). At the same time, OFA embeds a ConvLSTM (Xingjian et al. 2015) with \tanh function to OF_{fwd} to exploit information from consecutive US sequences and ensures the spatiotemporal smoothness. ConvLSTM not only establishes temporal relations like LSTM but also depicts local spatial features like CNN. Finally, OFa applies the *differentiable inverse warping* again to the \tilde{I} utilizing the predicted motion fields. OFA evaluates the similarity between the warped images and the manually segmented target labels. OFA simultaneously optimizes the segmentation twice and motion estimation once in OFA. The co-promotion of OFA makes OF-MSRN an efficient and reliable resolution for comprehensive analysis of carotid artery.

As a result, we get a composite loss function of OF-MSRN consisting of regression error, image similarity error, smoothness penalty and geometric consistency error: $L = \lambda_{Reg} L_{Reg} + \lambda_{Seg} L_{Sim} + \lambda_{OF} L_{OF} + \lambda_{OFA} L_{SWar}$, where λ_{Reg} (0.1), λ_{Seg} (0.01), λ_{OF} (1) and λ_{OFA} (1) are trade-off parameters for different tasks.

Experiments and Evaluation

Data Collection and Experiment Configuration

The proposed OF-MSRN has been intensively evaluated on a challenging dataset that including 101 clinical patients from the same center. Average years of patient age is 32 ± 9 with 58 males and 43 females, both left and right carotid arteries of these patients are examined. The 202 synthetic US sequences are generated by *MINDRAY Resona 7*. Each sequence has 175 frames with frame time is 28.57 ms. In each sequence, 45 consecutive frames containing at least one period are selected for the experiment. The ground truths are obtained manually according to the clinical criterion utilizing *ITK-SNAP* tool (Yushkevich et al. 2006) and our tool developed by *MATLAB*, and two carotid physicians with more than 10-year experiences double-check them. All images are uniformly cropped to 256×256 pixels. In the resized cropped image space, standard five-fold-cross-validation for

	Normal		Diastolic		Diseased	
	CALD	CIMT	CALD	CIMT	CALD	CIMT
OF-MSRN	94.34	8.11	101.96	9.39	91.27	9.54
MRN _{ns}	94.79	8.98	99.89	8.16	91.97	9.81
MSRN	93.95	7.32	99.38	7.85	90.74	9.03
MFE	91.26	6.98	95.47	11.82	87.43	8.55
IR-V4	87.73	5.33	92.09	6.79	84.04	7.91
RESNET	83.04	5.36	90.65	5.36	82.21	7.33
Ground Truth	96.00	9.00	102.00	9.00	93.00	11.00

Figure 5: Visualization of the automated direct quantitative measurement of the normal, diastolic, and diseased carotid artery. The table visualizes the results of six different methods for CALD and CIMT. The OF-MSRN achieves the most accurate prediction (Unit: pixel) of the diastolic and diseased carotid artery.

evaluation and comparison is employed to divide the training and testing sets. OF-MSRN is trained using adaptive moment estimation (Adam) optimizer with a moving average decay of 0.999, a shuffled batch size of 3. We use a decreasing learning rate with an initial value of $1e-4$. All the networks are based on *Python v3.6* and *TensorFlow v1.8* library, they run on *Intel (R) Core (TM) i7-7700 CPU @ 3.60GHz* with *12.0GB RAM* and *Nvidia GeForce GTX 1050 with cuDNN v7.3*.

Table 1: OF-MSRN achieves the lowest MAE in CALD (pixel) and CIMT (pixel) on the US dataset compared to relevant deep learning based methods.

	MAE_{CALD}	MAE_{CIMT}
OF-MSRN	1.23 ± 0.38	0.93 ± 0.22
MRN _{ns}	1.47 ± 0.41	1.05 ± 0.26
MSRN	2.38 ± 0.52	1.57 ± 0.35
MFE	5.21 ± 0.84	2.16 ± 0.49
IR-V4	8.65 ± 0.73	2.87 ± 0.41
RESNET	10.94 ± 0.81	3.27 ± 0.55

Result Analysis

Overall performance. We visualize and evaluate the prediction of OF-MSRN to prove its effectiveness and advantages. (1) The results of quantitative measurement are shown in Figure 5. Obviously, the diastolic and diseased NLI and FLI are very blurred, which is an exceedingly intractable challenge to achieve a comprehensive analysis of the carotid artery. OF-MSRN has overcome the challenge well and obtained accurate results (Figure 5). We evaluate the measurement effectiveness by the mean absolute error (MAE): $MAE = \frac{1}{M} \sum_{m=1}^M |p_m - l_m|$, where p_m and

Table 2: OF-MSRN achieves the highest IoU in CALD and CIMT on the US dataset compared to relevant deep learning based methods.

	IoU_{CALD}	IoU_{CIMT}
OF-MSRN	0.908 ± 0.008	0.893 ± 0.015
OF-MFE	0.838 ± 0.018	0.807 ± 0.045
MSRN	0.869 ± 0.013	0.874 ± 0.021
OF-FCN	0.799 ± 0.037	0.781 ± 0.049

l_m denote the prediction and the ground truth respectively, and M is the number of testing data. As shown in Table 1, OF-MSRN has low MAE (pixels) in both CALD and CIMT. (2) OF-MSRN has achieved excellent performance in the automated segmentation of CALD and CIMT (Figure 6). For normal carotid artery, the predicted boundaries are highly consistent with the ground truth. For diastolic carotid artery, OF-MSRN has greatly weakened the influence of motion artifact with the auxiliary of OFA, and the predicted boundaries are very close to the ground truth. For diseased carotid artery, OF-MSRN has achieved reliable prediction by taking advantage of dual optimization of the OFA. Similarly, we evaluation the Intersection over Union (IoU) of the constructed segmentation images in OFA. The IoU is the ratio of the overlapping area of ground truth and segmentation to the total area. IoU is defined as $IoU = \frac{Segmentation \cap GroundTruth}{Segmentation \cup GroundTruth}$. As shown in Table 2, OF-MSRN has achieved accurate segmentation of CALD and CIMT. (3) OF-MSRN is capable of predicting dense estimated motion fields. When taking the starting frame of the systolic motion as the target frame, OF-MSRN predicts a reliable motion field at each pixel (Figure 7). We visualize the results in Figure 7 (d) utilizing the color encoding. The color encodes flow direction and the color intensity encodes magnitude. Benefit from the Eq. (3) and Eq. (4), we effectively ensure the smoothness and consistency of the *differentiable inverse warping* (Figure 7 (c)). Therefore, OF-MSRN has great performance in comprehensive analysis for different subjects and pathologies of the carotid artery.

Performance comparison. The superiority of OF-MSRN for direct quantitative measurement and automated segmentation has been demonstrated by intra-comparisons and inter-comparisons. (1) We compare the measurement results of OF-MSRN with five methods (Figure 5 and Table 1). The results show that OF-MSRN has the most accurate prediction for both diastolic and diseased carotid arteries. **Intra-comparisons:** MRN_{ns}, a two-dimensional regression that uses MRN for measurement only; MSRN, a high dimensional regression that uses MRN for measurement and segmentation. MFE, a 2D regression based on MFE. MRN_{ns} performs better than OF-MSRN in the normal carotid artery which is not in radial motion, but it cannot overcome the interference of radial motion and diseased tissue. MRN_{ns} is a 2D regression that only for direct quantitative measurement. However, OF-MSRN has to predict extra boundary coordinates. MRN_{ns} is far less computationally expensive for solving the regression problem than OF-MSRN. The results of MSRN are far less accu-

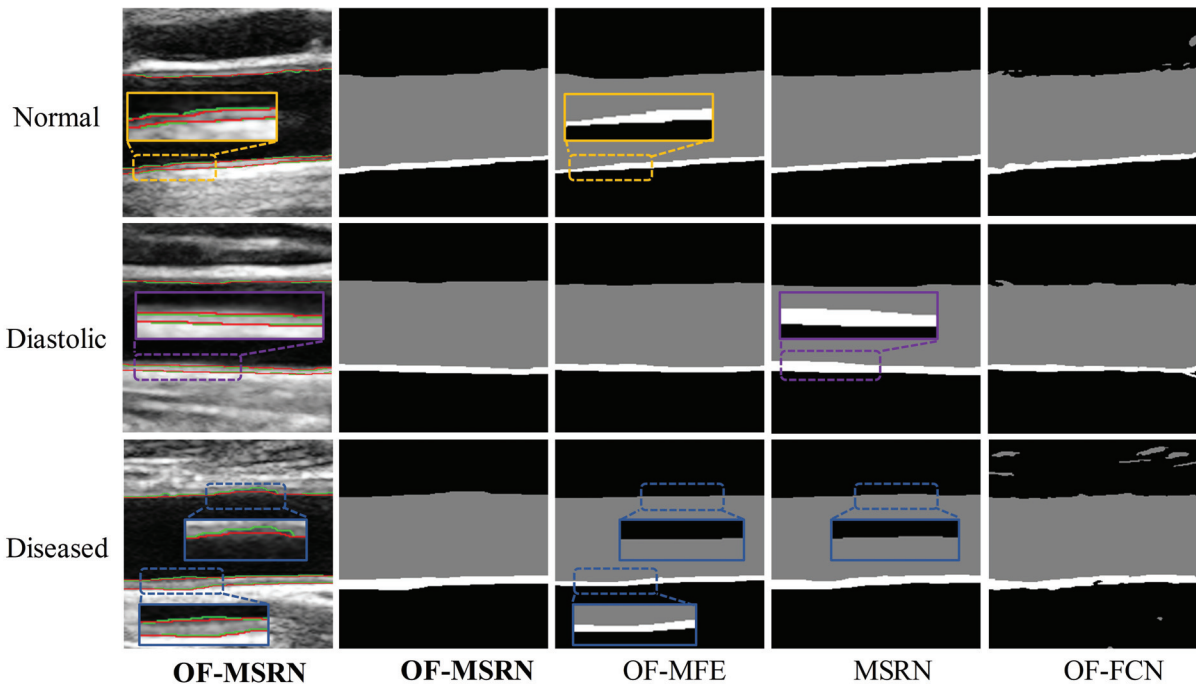


Figure 6: Visualization of the automated segmentation of the normal, diastolic, and diseased carotid artery. The red lines in the first column are the OF-MSRN predicted boundaries, and the green lines are the ground truth. The OF-MSRN has achieved the most accurate result compared to relevant deep learning based methods.

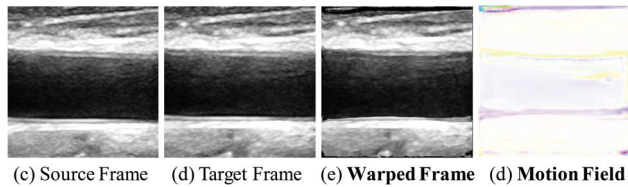


Figure 7: The OF-MSRN effectively predicts and visualizes the periodical motion of carotid artery. (a) Starting frame of the systole; (b) A frame in the systolic process; (c) Fake target frame warped by the *differentiable inverse warping*; (d) Motion field from the target frame to the source frame.

rate than that of OF-MSRN which is benefits from the OFA. It demonstrates that the improvement of segmentation results in the improvement of direct quantitative measurement. The results of MFE are especially low, which demonstrates that OF-MSRN has effectively extracted abundant structural information and discriminant features. **Inter-comparisons:** IR-V4 and RESNET denote the regression networks based on Inception-Resnet-V4 (Szegedy et al. 2017) and Resnet101 (He et al. 2016). The results show that both of them have poor performance and instability. (2) We compare the segmentation results of OF-MSRN with three methods (Figure 6 and Table 2). **Intra-comparisons:** OF-MFE, an OF-auxiliary MFE for segmentation. The results show that: 1) OF-MSRN extracts more detailed structural

information and segments the CIMT more accurately (yellow boxes in Figure 6) than OF-MFE; 2) OF-MSRN handles the artifact better than MSRN (purple boxes in Figure 6); 3) OF-MSRN achieves the most accurate segmentation of the diseased carotid artery (blue boxes in Figure 6). **Inter-comparisons:** OF-FCN, an OF-auxiliary FCN network for segmentation only. Even promoted by motion estimation, OF-FCN has achieved the worst performance. Therefore, OF-MSRN achieves superior performance that demonstrates its strengths in addressing the pathological correlations and structures variability.

Conclusion

We present a novel OF-auxiliary multi-task regression network named OF-MSRN for a comprehensive analysis of the carotid artery. The OF-MSRN extracts effective structural information by multi-scale feature concatenation and dual selection for simultaneous direct quantitative measurement and boundary segmentation. Meanwhile, the OF-MSRN uses an unsupervised bidirectional OF coder to estimate the periodic motion. More importantly, we creatively propose the OFA to construct segmentation images and achieve co-promotion of segmentation and motion estimation. Extensive experimental results have demonstrated the effectiveness and superiority of OF-MSRN. For future work, we will validate our network on multi-modal data of the carotid artery, and will also achieve better visualization of the dense motion field.

References

- Alan, S.; Ulgen, M. S.; Ozturk, O.; Alan, B.; Ozdemir, L.; and Toprak, N. 2003. Relation between coronary artery disease, risk factors and intima-media thickness of carotid artery, arterial distensibility, and stiffness index. *Angiology* 54(3):261–267.
- Baltgaile, G. 2012. Arterial wall dynamics. *Perspectives in Medicine* 1(1-12):146–151.
- Biswas, M.; Kuppili, V.; Araki, T.; Edla, D. R.; Godia, E. C.; Saba, L.; Suri, H. S.; Omerzu, T.; Laird, J. R.; Khanna, N. N.; et al. 2018. Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on japanese diabetic cohort. *Computers in biology and medicine* 98:100–117.
- Biswas, M.; Kuppili, V.; Saba, L.; Edla, D. R.; Suri, H. S.; Sharma, A.; Cuadrado-Godia, E.; Laird, J. R.; Nicolaidis, A.; and Suri, J. S. 2019. Deep learning fully convolution network for lumen characterization in diabetic patients using carotid ultrasound: a tool for stroke risk. *Medical & biological engineering & computing* 57(2):543–564.
- Bots, M. L.; Grobbee, D. E.; Hofman, A.; and Witteman, J. C. 2005. Common carotid intima-media thickness and risk of acute myocardial infarction: the role of lumen diameter. *Stroke* 36(4):762–767.
- Bots, M. L.; Hofman, A.; and Grobbee, D. E. 1997. Increased common carotid intima-media thickness: adaptive response or a reflection of atherosclerosis? findings from the rotterdam study. *Stroke* 28(12):2442–2447.
- Cheng, J.; Tsai, Y.-H.; Wang, S.; and Yang, M.-H. 2017. Segflow: Joint learning for video object segmentation and optical flow. In *Proceedings of the IEEE international conference on computer vision*, 686–695.
- Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; and Brox, T. 2015. FlowNet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 2758–2766.
- Golemati, S.; Stoitsis, J. S.; Gastouniotti, A.; Dimopoulos, A. C.; Koropouli, V.; and Nikita, K. S. 2012. Comparison of block matching and differential methods for motion analysis of the carotid artery wall from ultrasound images. *IEEE Transactions on Information Technology in Biomedicine* 16(5):852–858.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; and Brox, T. 2017. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2462–2470.
- Jaderberg, M.; Simonyan, K.; Zisserman, A.; and Kavukcuoglu, K. 2015. Spatial transformer networks. In *Advances in neural information processing systems*, 2017–2025.
- Naik, V.; Gamad, R.; and Bansod, P. 2013. Carotid artery segmentation in ultrasound images and measurement of intima-media thickness. *BioMed research international* 2013.
- Naqvi, T. Z., and Lee, M.-S. 2014. Carotid intima-media thickness and plaque in cardiovascular risk assessment. *JACC: Cardiovascular Imaging* 7(10):1025–1038.
- Qin, C.; Bai, W.; Schlemper, J.; Petersen, S. E.; Piechnik, S. K.; Neubauer, S.; and Rueckert, D. 2018. Joint learning of motion estimation and segmentation for cardiac mr image sequences. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 472–480. Springer.
- Shin, J.; Tajbakhsh, N.; Todd Hurst, R.; Kendall, C. B.; and Liang, J. 2016. Automating carotid intima-media thickness video interpretation with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2526–2535.
- Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8934–8943.
- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; and Alemi, A. A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- Tsai, Y.-H.; Yang, M.-H.; and Black, M. J. 2016. Video segmentation via object flow. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3899–3908.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P.; et al. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13(4):600–612.
- Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; and He, K. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500.
- Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, 802–810.
- Yin, Z., and Shi, J. 2018. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1983–1992.
- Yushkevich, P. A.; Piven, J.; Hazlett, H. C.; Smith, R. G.; Ho, S.; Gee, J. C.; and Gerig, G. 2006. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31(3):1116–1128.