Weakly-Supervised Coronary Artery Segmentation from DSA Sequence via Motion-Aware Modeling

Han Wu[1,2,4], Xiaosong Xiong[1], Yanli Song[2], Yiqiang Zhan[2], Sean Zhou[2], Dijia Wu[2(✉)], and Dinggang Shen[1,2,3(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[2] Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China
[3] Shanghai Clinical Research and Trial Center, Shanghai, China
[4] Lingang Laboratory, Shanghai, China
dijia.wu@uii-ai.com, dinggang.shen@gmail.com

Abstract. Accurate coronary artery segmentation from Digital Subtraction Angiography (DSA) is critical for computer-aided cardiovascular diagnosis and intervention. In DSA, contrast wash-in provides characteristic temporal perfusion cues, i.e., vessels progressively enhance over time, which can help distinguish true vessels from dark non-vessel tubular structures and improve recognition under low-contrast conditions by comparing neighboring frames. However, exploiting such cues is non-trivial, as cardiac motion introduces non-rigid vessel deformation and misalignment across frames. Clinically, radiologists typically review the entire dynamic injection process, but measure only a single key frame where the vessel is maximally opacified. Based on this observation, we present **MACOS**, a **M**otion-**A**ware **CO**ronary **S**egmentation framework that leverages long-range temporal injection context to supervise segmentation only on a key frame with maximal opacification. Specifically, we first propose a Motion-Aware Gated Recurrent Unit (GRU) module to explicitly estimate inter-frame displacement fields to align features before aggregation, thus enabling robust spatiotemporal modeling under cardiac motion. Then, to leverage unlabeled intermediate frames, we further introduce a Physics-Informed Vessel Loss to enforce monotonic vessel growth driven by contrast wash-in, for providing effective regularization without requiring dense annotations. Experiments on 971 DSA sequences demonstrate that MACOS significantly outperforms state-of-the-art methods, improving robustness on typical challenging cases. To the best of our knowledge, MACOS is the first framework to exploit long coronary DSA sequences for weakly-supervised vessel segmentation with only key-frame annotation, and is validated on the largest coronary DSA video dataset to date. Code is available at https://github.com/Fitz-Fitz/MACOS.

Keywords: Coronary Segmentation · Digital Subtraction Angiography · Weakly-Supervised Learning · Physics-Informed Constraints

1 Introduction

Digital Subtraction Angiography (DSA) serves as the gold standard for cardiovascular diagnosis and interventional planning [4,14], where accurate segmentation of coronary vessels is crucial for visualizing vascular structures and computing quantitative clinical metrics [12]. In standard clinical workflows, radiologists typically review the entire dynamic injection process but measure only the key frame where the vessel is fully opacified.

Existing methods primarily operate on these key frames and fall into two paradigms: (1) Single-frame models [15,8,3] that process the key frame independently using FCN-based architectures [15] or multi-scale attention [3] to exploit the vessel structure from static frame; (2) Sequence-based models [5,6] that incorporate a few preceding frames for local temporal context, i.e., employing 3D convolutions in SVS-Net [5] and TVS-Net [6] for spatiotemporal feature extraction. However, both paradigms have critical limitations. Single-frame methods ignore temporal perfusion cues induced by contrast wash-in, which are informative for distinguishing true vessels from dark non-vessel tubular structures and improving recognition under low contrast. Sequence-based methods suffer from: (i) Limited temporal scope, i.e., using only 2–4 frames (0.1–0.3 seconds) near key frames [5], thus missing the onset-to-peak wash-in process whose temporal perfusion cues can provide a strong reference for suppressing static tubular artifacts and recovering low-contrast segments; (ii) Motion-agnostic modeling, i.e., simply using 3D convs by assuming spatial alignment across frames without consideration of non-rigid cardiac deformation, thus leading to feature blurring.

To address these limitations, we present **MACOS** (**M**otion-**A**ware **CO**ronary **S**egmentation), the first framework to exploit long coronary DSA sequences with explicit motion compensation under weakly-supervised learning. To summarize, our main contributions are: (1) We propose a Motion-Aware (MA) GRU module to explicitly estimate displacement fields to align temporal features before aggregation, enabling robust spatiotemporal modeling under cardiac motion. (2) We introduce a Global Motion Loss and a Physics-Informed Vessel Loss to leverage the hemodynamic monotonicity of contrast wash-in to regularize intermediate features, enabling effective weakly-supervised learning with only key frame annotation. (3) We validate MACOS on the largest coronary DSA video dataset to date (971 long sequences), outperforming SOTA methods, and also systematically analyze the impact of temporal configuration.

2 Methodology

2.1 Problem Formulation

Let $\mathbf{V} = \{I_1, I_2, \ldots, I_t, \ldots, I_T\}$ denote a DSA sequence of T frames, where $I_t \in \mathbb{R}^{H \times W}$ represents the t-th frame. Following clinical practice, the last frame I_T corresponds to the key frame with maximal vessel opacification, for which a binary segmentation mask $Y \in \{0,1\}^{H \times W}$ is provided. Our goal is to learn f_θ that predicts the vessel map $\hat{P} = f_\theta(\mathbf{V})$ for the key frame by exploiting

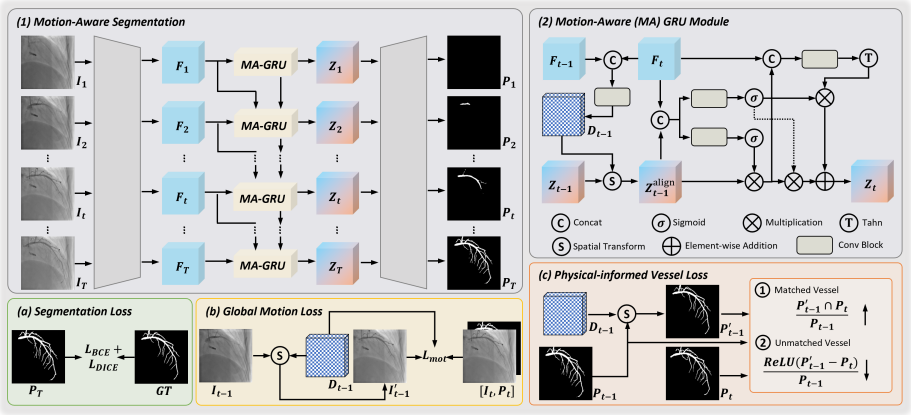

Fig. 1: An overview of the proposed MACOS framework.

spatiotemporal coherence across the entire sequence. We operate in a weakly-supervised setting: only the key frame is annotated, while unlabeled intermediate frames $\{I_t\}_{t<T}$ are used to provide temporal perfusion cues from contrast wash-in and context for handling ambiguities under cardiac motion.

2.2 Overview

Fig. 1 illustrates the overall architecture of MACOS. Given an input DSA sequence $\mathbf{V} = \{I_1, I_2, \ldots, I_t, \ldots, I_T\}$, we first employ a shared U-Net-based encoder to extract hierarchical spatial features independently for each frame:

$$\mathbf{F}_t = f_\theta^{enc}(I_t) = \{F_t^1, F_t^2, \ldots, F_t^L\}, \quad t \in [1, T], \tag{1}$$

where $F_t^l \in \mathbb{R}^{C_l \times H_l \times W_l}$ denotes the feature map at scale l with spatial resolution downsampled by a factor of 2^{l-1} of frame I_t and $l \in [1, L]$ denotes the scale index with totally L scales.

To capture long-range temporal perfusion dynamics while explicitly handling cardiac motion, we propose a MA-GRU module operating at each scale l. At time t, it takes current frame feature F_t^l, previous frame feature F_{t-1}^l, and previous hidden state Z_{t-1}^l, outputting the updated state and displacement:

$$Z_t^l, \mathcal{D}_{t-1}^l = \text{MA-GRU}(F_t^l, F_{t-1}^l, Z_{t-1}^l), \tag{2}$$

where $\mathcal{D}_{t-1}^l \in \mathbb{R}^{2 \times H_l \times W_l}$ represents the estimated displacement field from frame $t-1$ to frame t at scale l. Critically, Z_{t-1}^l is spatially warped by $\mathcal{D}_{t-1}^l$ to align with the current frame before applying GRU gating operations, ensuring motion-compensated temporal aggregation which will be detailed in Sec. 2.3.

The motion-aligned spatiotemporal features $\{Z_t^l\}_{l=1}^L$ are fed into a shared decoder that progressively upsamples and fuses multi-scale features to reconstruct

high-resolution features G_t^l:

$$G_t^l = \mathcal{F}_{dec}(\mathrm{Up}(G_t^{l+1}) \oplus Z_t^l), \quad l = L-1, \ldots, 1, \tag{3}$$

where $G_t^L = Z_t^L$, $\oplus$ denotes channel-wise concatenation, and $\mathcal{F}_{dec}$ consists of convolutional blocks. A 1×1 convolution head on G_t^1 produces the per-frame segmentation map $P_t \in \mathbb{R}^{H \times W}$. This generates a sequence of predictions $\mathbf{P} = \{P_1, \ldots, P_T\}$, where P_T is the final output for the key frame with direct supervision, while intermediate predictions $\{P_t\}_{t<T}$ are regularized through our physics-informed loss (detailed in Sec. 2.4). The segmentation loss for the key frame combines binary cross-entropy and Dice:

$$\mathcal{L}_{seg} = \mathcal{L}_{BCE}(P_T, Y) + \mathcal{L}_{Dice}(P_T, Y). \tag{4}$$

2.3 Motion-Aware GRU Module

Standard ConvGRU assumes spatial correspondence between frames, without consideration of cardiac motion, which causes blurred features and ghosting artifacts that degrade segmentation quality. Our MA-GRU (Fig. 1(b)) explicitly estimates and compensates inter-frame motions before aggregation. Given the current spatial feature F_t^l and previous feature F_{t-1}^l at scale l, a lightweight motion estimation network predicts the inter-frame displacement field:

$$\mathcal{D}_{t-1}^l = \mathcal{M}_\phi([F_t^l, F_{t-1}^l]), \tag{5}$$

where $\mathcal{M}_\phi$ consists of three convolutional layers with kernel size 3×3, and $[\cdot, \cdot]$ denotes channel-wise concatenation. The estimated displacement field is then used to warp the previous hidden state to align it with the current frame:

$$Z_{t-1}^{l,\mathrm{align}} = \mathcal{W}(Z_{t-1}^l, \mathcal{D}_{t-1}^l), \tag{6}$$

where $\mathcal{W}(\cdot, \cdot)$ denotes bilinear spatial warping using a differentiable grid sampler. The motion-aligned hidden state $Z_{t-1}^{l,\mathrm{align}}$ is then used as:

$$\begin{aligned} U_t^l &= \sigma(\mathrm{Conv}([F_t^l, Z_{t-1}^{l,\mathrm{align}}])), \quad R_t^l = \sigma(\mathrm{Conv}([F_t^l, Z_{t-1}^{l,\mathrm{align}}])), \\ \tilde{Z}_t^l &= \tanh(\mathrm{Conv}([F_t^l, R_t^l \odot Z_{t-1}^{l,\mathrm{align}}])), \\ Z_t^l &= (1 - U_t^l) \odot \tilde{Z}_t^l + U_t^l \odot Z_{t-1}^{l,\mathrm{align}}, \end{aligned} \tag{7}$$

where U_t^l and R_t^l are the *update* and *reset* gates, respectively, σ is the sigmoid activation, and $\odot$ denotes element-wise multiplication. By operating on motion-aligned features, this formulation ensures temporal aggregation to focus on anatomically consistent vascular structures across frames, thus significantly enhancing robustness against cardiac motion artifacts.

To supervise the motion estimation, we introduce the Global Motion Loss $\mathcal{L}_{mot}$, which implements masked photometric consistency. Unlike natural videos, DSA sequences violate the brightness constancy assumption due to progressive

contrast agent wash-in, where vessel intensities increase over time. We therefore restrict photometric supervision to background motion by masking out vessel areas:

$$\mathcal{L}_{mot} = \frac{1}{T-1}\sum_{t=2}^{T}\left(\|(1-P_t)\odot(I_t - \mathcal{W}(I_{t-1}, \mathcal{D}^1_{t-1}))\|_1 + \gamma\|\nabla\mathcal{D}^1_{t-1}\|_2^2\right), \quad (8)$$

where P_t is the predicted vessel map that identifies contrast-enhanced regions to be excluded, and the spatial gradient term $\gamma\|\nabla\mathcal{D}^1_{t-1}\|_2^2$ enforces smoothness for anatomically plausible motion fields. Complementarily, the displacement in vessel regions is supervised through physics-informed constraints that exploit hemodynamic monotonicity, as detailed in Sec. 2.4.

2.4 Physics-Informed Vessel Loss

During contrast injection, vascular structures progressively emerge from invisible to fully opacified, exhibiting monotonic vessel growth. That is, visible regions only expand or persist, and never shrink, until peak opacification. Unlike photometric consistency that relies on brightness constancy (violated by contrast wash-in), this monotonicity constraint provides a physics-based prior that holds specifically in vessel regions, enabling spatiotemporal regularization without any annotation.

Formally, let $P_t \in \mathbb{R}^{H\times W}$ denote the predicted vessel map at time t. We warp P_{t-1} forward to frame t using the estimated displacement field as $P'_{t-1} = \mathcal{W}(P_{t-1}, \mathcal{D}^1_{t-1})$, where $\mathcal{D}^1_{t-1}$ is the finest-scale displacement field from Sec. 2.3. Under monotonic growth, the motion-compensated response should be non-decreasing, i.e., $P'_{t-1}(i,j) \in P_t(i,j)$ for all pixels (i,j). We thus penalize the fraction of vessel that disappears after alignment:

$$\begin{aligned}\mathcal{L}_{phy}(t) &= \frac{\sum_{i,j}\mathrm{ReLU}(P'_{t-1}(i,j) - P_t(i,j))}{\sum_{i,j}P'_{t-1}(i,j)+\epsilon} \\ &= 1 - \frac{\sum_{i,j}\min(P'_{t-1}(i,j), P_t(i,j))}{\sum_{i,j}P'_{t-1}(i,j)+\epsilon},\end{aligned} \quad (9)$$

The numerator accumulates the *unmatched* vessel response $\mathrm{ReLU}(P'_{t-1} - P_t)$, normalized by the total vessel in P'_{t-1}. Equivalently, minimizing this unmatched vessel maximizes the *matched* overlap vessel $\min(P'_{t-1}, P_t)$. $\epsilon = 10^{-6}$ prevents division by zero. The Physics-Informed Vessel Loss averages over all temporal transitions:

$$\mathcal{L}_{phy} = \frac{1}{T-1}\sum_{t=2}^{T}\mathcal{L}_{phy}(t). \quad (10)$$

By propagating constraints backward from the fully-supervised key frame through intermediate predictions, this loss provides effective regularization for temporal consistency without requiring dense frame-level annotation, enabling scalable weakly-supervised learning.

2.5 Overall Training Objective

The complete training objective combines three complementary loss terms:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_1 \mathcal{L}_{mot} + \lambda_2 \mathcal{L}_{phy}, \tag{11}$$

where $\mathcal{L}_{mot}$ denotes the Global Motion Loss, $\mathcal{L}_{phy}$ denotes the Physics-Informed Vessel Loss, and the hyperparameters λ_1 and λ_2 balance these objectives.

3 Experiments

3.1 Dataset and Evaluation Metrics

Dataset. We evaluate MACOS on a large-scale coronary DSA dataset collected from two clinical centers, comprising 971 angiography sequences with a duration of 3 seconds each. Each sequence is standardized such that the first frame captures the pre-injection phase (with no visible contrast), and the last frame corresponds to the key frame with maximal vessel opacification. Ground-truth vessel masks for key frames were manually delineated by three experienced interventional radiologists to ensure annotation quality. The dataset is partitioned at the patient level into 703/72/196 for training, validation and testing, respectively. It provides the first long-sequence benchmark and the largest dataset for coronary DSA vessel segmentation to date. This represents a significant advancement over existing sequence-based datasets [5,6], which contain only 2–4 frames near to the key frame spanning 0.1–0.3 seconds (120 and 60 sequences, respectively).

Evaluation Metrics. We evaluate segmentation performance using five standard metrics: Dice score (Dice), Intersection over Union (IoU), Precision, Recall, and Average Surface Distance (ASD in pixels).

3.2 Implementation Details

MACOS is implemented in PyTorch and trained on NVIDIA H20 GPU with a batch size of 8. The backbone adopts a standard U-Net [9] architecture, processing inputs at resolution 256×256. For each training sequence, we uniformly sample $T = 6$ frames spanning the entire injection process. The network is trained for 150 epochs using the Adam optimizer with an initial learning rate of 10^{-4}, which is reduced by a factor of 0.1 at epochs 50 and 100. Loss balancing weights are set to $\lambda_1 = 0.3$ and $\lambda_2 = 0.5$ based on validation performance. Data augmentation strategies include random rotation ($\pm 15^\circ$), horizontal/vertical flipping, and Gaussian noise to improve generalization.

Table 1: Comparison with SOTA methods. Values are presented as mean (std).

Method	Dice (%)	IoU (%)	Precision (%)	Recall (%)	ASD (pixel)
UNet [9]	80.18(6.32)	67.48(8.15)	81.63(7.12)	79.84(7.45)	2.94(1.99)
AttentionUNet [11]	81.14(5.98)	69.49(7.88)	81.81(6.54)	80.50(7.11)	2.78(1.63)
UNet++ [16]	81.58(5.45)	69.47(7.21)	81.76(6.12)	82.36(6.89)	2.68(1.55)
SwinUNet [1]	80.99(6.14)	67.94(7.94)	80.53(6.87)	81.39(7.23)	3.09(1.92)
TransUNet [2]	81.80(5.87)	68.44(7.76)	81.19(6.65)	82.22(7.04)	2.99(1.81)
VM-UNet [10]	81.27(5.79)	69.06(7.65)	82.51(6.23)	80.97(7.15)	2.81(1.75)
HVM-UNet [13]	82.56(5.23)	71.70(7.15)	83.11(5.95)	81.24(6.67)	2.41(1.45)
nnUNet [7]	<u>84.07</u>(4.92)	<u>73.31</u>(7.05)	<u>85.17</u>(5.88)	82.22(6.54)	<u>1.77</u>(1.02)
SVS-Net [5]	83.17(5.40)	71.81(7.10)	82.32(6.01)	83.40(6.32)	1.88(1.15)
TVS-Net [6]	83.99(5.25)	72.76(6.95)	84.83(5.82)	<u>83.99</u>(6.15)	1.96(1.21)
MACOS	**85.47**(4.67)	**74.90**(6.83)	**86.55**(5.77)	**84.90**(6.75)	**1.47**(0.92)

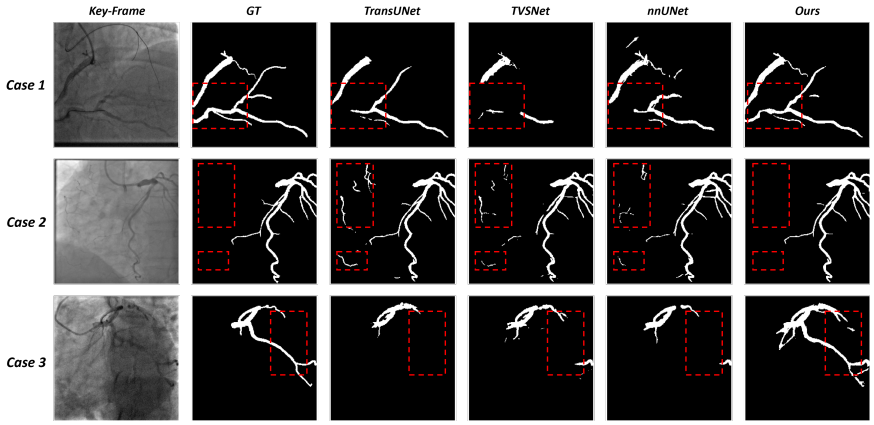

Fig. 2: Visual comparison on three typical challenging cases including out of plane motion (Case 1), low-contrast capillaries prone to false negatives (Case 2), and boundary ambiguities from complex background interference (Case 3).

3.3 Comparison with SOTA Approaches

We compare MACOS against state-of-the-art methods, including CNN-based (UNet [9], Attention UNet [11], UNet++ [16], nnUNet [7]), Transformer-based (SwinUNet [1], TransUNet [2]), Mamba-based (VM-UNet [10], HVM-UNet [13]), and sequence-based (SVS-Net [5], TVS-Net [6]). For fair comparison, single-frame methods are trained on key frames only, while sequence-based methods use the same temporal input as MACOS.

As shown in Table 1, MACOS achieves the best performance across all metrics. Compared with the second best nnUNet [7], MACOS improves Dice by 1.40%, demonstrating the benefit of leveraging temporal information. Compared to sequence-based methods SVS-Net [5] and TVS-Net [6] with 3D convolutional aggregation, MACOS highlights its advantage of explicit motion compensation

Table 2: Ablation on component contributions and temporal configurations.

Configuration	Dice (%)	IoU (%)	Precision (%)	Recall (%)	ASD (pixel)
Component Analysis					
Baseline (U-Net)	80.18(6.32)	67.48(8.15)	81.63(7.12)	79.84(7.45)	2.94(1.99)
+ ConvGRU	83.06(5.40)	71.47(7.40)	84.02(6.30)	82.76(6.90)	2.05(1.45)
+ MA-GRU	84.76(4.90)	73.34(6.95)	85.44(6.00)	83.92(6.60)	1.65(1.10)
+ Physics Loss (Full)	**85.47**(4.67)	**74.90**(6.83)	**86.55**(5.77)	**84.90**(6.75)	**1.47**(0.92)
Temporal Configuration (Frame Rate)					
MACOS (FPS=1)	85.00(4.75)	74.25(6.95)	86.05(5.95)	84.55(6.90)	1.55(0.98)
MACOS (FPS=2)	**85.47**(4.67)	**74.90**(6.83)	**86.55**(5.77)	**84.90**(6.75)	**1.47**(0.92)
MACOS (FPS=3)	85.28(4.70)	74.60(6.85)	86.32(5.85)	84.75(6.80)	1.50(0.95)
MACOS (FPS=4)	85.14(4.72)	74.40(6.90)	86.18(5.90)	84.63(6.85)	1.52(0.96)

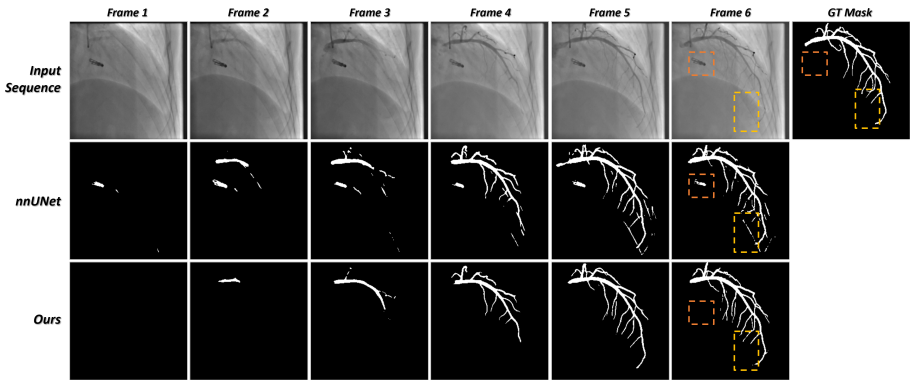

Fig. 3: Visualization of temporal prediction on the whole sequence.

for long-sequence modeling. MACOS also achieves much better segmentation performance on three typical challenging cases as presented in Fig. 2. Fig. 3 further shows temporally visualization across the injection sequence, finally leading to a more accurate segmentation on the key frame.

3.4 Ablation Studies

Component Analysis. As shown in Table 2 (top), we ablate key components of MACOS. The baseline U-Net achieves 80.18% Dice. Adding ConvGRU improves performance to 83.06% Dice, validating the benefit of temporal aggregation. Replacing ConvGRU with MA-GRU further improves Dice to 84.76%, demonstrating that explicit motion compensation better handles cardiac motion. Finally, incorporating the Physics-Informed Vessel Loss yields the best performance while reducing ASD, confirming that hemodynamic constraints can provide effective regularization without requiring additional annotation.

Impact of Temporal Configuration. We investigate how frame rate affects segmentation. As shown in Table 2 (bottom), lower FPS degrades performance as large inter-frame motion exceeds the compensation capacity, while higher FPS yields smoother motion but diminishing returns due to redundant frames under sparse supervision. The optimal trade-off is achieved at FPS=2.

4 Conclusion

We present MACOS, which to the best of our knowledge is the first motion-aware framework to exploit long coronary DSA sequences for weakly-supervised vessel segmentation with only key-frame supervision. Experiments on 971 clinical DSA sequences (the largest coronary DSA video dataset to date) show that MACOS achieves state-of-the-art performance and improves robustness under typical challenging conditions, highlighting the value of explicit motion compensation and physiology-guided temporal regularization.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers U23A20295, 62131015, 82441023, 82394432), the China Ministry of Science and Technology (STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article

References

1. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
2. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
3. Deng, H., Fang, T., Min, X.: Multi-scale across attention incorporated network for x-ray coronary vessel segmentation. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
4. Doenst, T., Haverich, A., Serruys, P., Bonow, R.O., Kappetein, P., Falk, V., Velazquez, E., Diegeler, A., Sigusch, H.: Pci and cabg for treating stable coronary artery disease: Jacc review topic of the week. Journal of the American College of Cardiology **73**(8), 964–976 (2019)
5. Hao, D., Ding, S., Qiu, L., Lv, Y., Fei, B., Zhu, Y., Qin, B.: Sequential vessel segmentation via deep channel attention network. Neural Networks **128**, 172–187 (2020)

6. He, H., Banerjee, A., Choudhury, R.P., Grau, V.: Deep learning based coronary vessels segmentation in x-ray angiography using temporal information. Medical Image Analysis **102**, 103496 (2025)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Jiang, Z., Ou, C., Qian, Y., Rehan, R., Yong, A.: Coronary vessel segmentation using multiresolution and multiscale deep learning. Informatics in Medicine Unlocked **24**, 100602 (2021)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
10. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications (2024)
11. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis **53**, 197–207 (2019)
12. Tu, S., Barbato, E., Köszegi, Z., Yang, J., Sun, Z., Holm, N.R., Tar, B., Li, Y., Rusinaru, D., Wijns, W., et al.: Fractional flow reserve calculation from 3-dimensional quantitative coronary angiography and timi frame count: a fast computer model to quantify the functional significance of moderately obstructed coronary arteries. JACC: Cardiovascular Interventions **7**(7), 768–777 (2014)
13. Wu, R., Liu, Y., Liang, P., Chang, Q.: H-vmunet: High-order vision mamba unet for medical image segmentation. Neurocomputing **624**, 129447 (2025)
14. Xiong, X., Jiang, C., Wu, P., Zhang, X., Song, Y., Zhang, X., Tao, Z., Wu, D., Shen, D.: Vascular-topology-aware deep structure matching for 2d dsa and 3d cta rigid registration. In: International Conference on Information Processing in Medical Imaging. pp. 94–107. Springer (2025)
15. Yang, S., Kweon, J., Roh, J.H., Lee, J.H., Kang, H., Park, L.J., Kim, D.J., Yang, H., Hur, J., Kang, D.Y., et al.: Deep learning segmentation of major vessels in x-ray coronary angiography. Scientific reports **9**(1), 16897 (2019)
16. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)