This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

# Association of Cardiovascular Risk Factors with Segmentation-Free Motion-Derived Coronary Artery Dynamics

Yasamin Moghbelan[1][0009−0005−2072−4385], Daniel Breucker[2,3,4][0009−0003−9200−565X], Moritz von Scheidt[2,4][0000−0001−7159−8271], Oliver Hayden[3][0000−0002−2678−8663], Javier Olivan Bescos[5], and Danny Ruijters[1,5][0000−0002−9931−4047]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
{y.moghbelan, d.ruijters}@tue.nl
[2] Department of Cardiology, TUM University Hospital German Heart Center, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany
{moritz.scheidt}@tum.de
[3] Heinz Nixdorf Chair of Biomedical Electronics, TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Germany
{daniel.breucker, oliver.hayden}@tum.de
[4] Deutsches Zentrum für Herz- und Kreislaufforschung (DZHK), Partner Site Munich Heart Alliance, Munich, Germany
[5] Philips Healthcare, Best, The Netherlands
{danny.ruijters, javier.olivan.bescos}@philips.com

**Abstract.** Coronary artery motion observed in X-ray angiography reflects complex physiological processes such as myocardial contraction, vessel compliance, and hemodynamic forces, yet it remains underused as a quantitative signal of cardiac dynamics because extracting it at scale from routine clinical data is technically demanding. This paper presents a scalable, segmentation-free, data-driven deep learning framework for large-scale analysis of coronary artery motion based on dense optical flow. Motion is extracted directly from routine angiographic sequences using Frangi-based vessel enhancement and deep optical flow estimation. Lacking ground-truth motion annotations, we adopt a dual validation strategy: synthetic data with exactly known flow, and real sequences assessed through reconstruction-based metrics and qualitative visualization. The extracted flow fields drive a video-based deep neural network that learns motion representations, evaluated for association with clinical risk factors using patient-level stratified train/validation splitting to prevent data leakage. From a cohort of 1,824 patients, projection-angle filtering and clinical label availability yielded analysis cohorts ranging from 124–316 patients (288–757 sequences) per parameter. Learned motion-derived representations predicted several cardiovascular risk factors above majority-class baselines, including ejection fraction, cholesterol level, hypertension, and nicotine use. These results position coronary artery motion from routine angiography as a scalable, clinically relevant signal warranting further investigation as a cardiovascular risk-association tool.

**Keywords:** Deep Learning · Vision · Coronary Artery Motion · Optical Flow · X-ray Angiography · Representation Learning · Segmentation-free · Cardiovascular Risk Factors · Motion-derived Biomarker.

## 1 Introduction

**Motivation and Contributions** Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for nearly one-third of all deaths [1], and a substantial proportion of patients with coronary artery disease present without established traditional risk factors [2]. This highlights the need for alternative functional biomarkers derived non-invasively from routine clinical data. Coronary artery (CA) motion in X-ray coronary angiography (XCA) encodes rich physiological information, reflecting stenosis severity, vessel compliance, myocardial contraction, and hemodynamic forces; and can in principle be obtained directly from routine sequences without additional imaging or manual vessel segmentation, making it well suited to large-scale analysis. Automatic extraction and interpretation of CA motion from routine XCA nonetheless remains challenging: vessels are thin and low-contrast; motion is non-rigid and confounded by respiration; acquisition geometry and contrast dynamics vary widely; and ground-truth motion annotations are absent. Existing approaches rely on manual feature extraction or are evaluated only on major vessels in small datasets, so a robust, data-driven, and segmentation-free framework for large-scale coronary motion analysis is still lacking.

We address this gap as an application study. Rather than proposing a new optical flow algorithm, we investigate whether automatically extracted, segmentation-free coronary motion representations are associated with cardiovascular risk factors, using a scalable analysis pipeline. Our contributions are:

1. A scalable, segmentation-free pipeline for extracting dense coronary artery motion from routine angiographic sequences, combining Frangi vessel enhancement with fine-tuned deep learning optical flow estimation.
2. A dual evaluation strategy for optical flow quality that combines synthetic data with known ground-truth motion and reconstruction-based quantitative assessment on real XCA sequences.
3. A clinical association study showing that learned coronary motion representations exhibit measurable associations with multiple cardiovascular risk factors.

**Related Work** Cardiac motion analysis has been studied across multiple imaging modalities. In coronary CT angiography (CCTA), motion is primarily treated as an artifact, with most work focusing on motion correction rather than motion interpretation [3], [4]. Magnetic Resonance Imaging (MRI)-based techniques such as tagged and cine MRI, enable quantification of myocardial deformation [5], while echocardiography similarly emphasizes myocardial strain and cardiac dynamics [6]. In these modalities, however, the geometric dynamics of coronary

arteries are not directly analyzed. Systematic analysis of CA motion has been explored mainly in XCA. Early studies by [7], [8] and [9] relied on handcrafted geometric descriptors and manual classification of selected vessel segments to study relationships with stenosis severity or biomechanical stress. These methods are insightful but labor intensive, limited to small datasets, and focused on major coronary branches. More recently, Yin et al. [10] operate on selected frame pairs using segmentation masks as supervision and evaluate on a small synthetic dataset. In contrast, our framework processes complete sequences of 12–60 frames without segmentation labels, evaluated on a single-center dataset of over 50,000 sequences from 1,824 patients. Rather than developing a new flow estimator, we investigate what clinically relevant information coronary motion carries at scale, using a fine-tuned pretrained flow model as a reliable motion-extraction tool within a clinical association study.

## 2 Methods

### 2.1 Data

**Dataset Description and Clinical Parameters:** We analyze a large angiography dataset from TUM University Hospital German Heart Center, comprising more than 50,000 XCA sequences from 1,824 patients acquired over ten years, with patients undergoing multiple examinations across sessions months apart. All sequences are grayscale at $512 \times 512$ pixels and contain 12–60 frames covering the contrast injection and washout phases. From 157 available clinical parameters, five were selected based on known association with coronary artery disease and broad availability: ejection fraction, cholesterol, diabetes, hypertension, and nicotine use.

**Data Distribution and Selection:** XCA sequences were acquired across a wide range of projection angles. To reduce geometric variability, we restrict analysis to the largest coherent cluster centered at $[-4°, 39°]$ in primary and secondary angles with a hard boundary of $12°$, yielding 8,792 sequences. From these, DICOM files were manually screened using Frangi-based vessel enhancement [11] and visual inspection to exclude cases with severe table motion, insufficient contrast injection, or large occlusions, any of which can cause vessels to partially or completely leave the field of view. Most of this reduction comes from quality screening rather than missing clinical labels; the per-field counts in Table 1 vary because not all patients had every parameter recorded. For each retained sequence, a fixed window of 9 consecutive frames is selected by maximizing nonzero Frangi vesselness over time, spanning injection through washout; from these, 8 optical flow fields are computed between consecutive pairs, giving a per-clip input tensor of shape $8 \times 2 \times 512 \times 512$.

**Data Splitting and Validation Protocol:** All train/validation splits were performed at the patient level. First, sequences were grouped by anonymized patient identifier. Patients, rather than individual sequences, were assigned to training or validation sets using an 80:20 stratified split. All sequences belonging to a given patient were then assigned exclusively to either the training or

validation set. We verified the absence of patient leakage by confirming that the intersection between training and validation patient identifiers was empty for each clinical parameter. Class imbalance was addressed using weighted binary cross-entropy for binary parameters and weighted cross-entropy for the three-class cholesterol task, together with augmentation via horizontal flipping ($p = 0.5$) and rotation ($\pm 15°$).

Table 1: Per-field dataset statistics and class distributions.

| Field | Patients | Sequences | Class Distribution |
|---|---|---|---|
| Ejection Fraction | 244 | 581 | C0 (EF<60%): 195 (33.6%), C1 (EF≥60%): 386 (66.4%) |
| Cholesterol | 316 | 757 | C0 (normal): 188 (24.8%), C1 (intermediate): 252 (33.3%), C2 (elevated): 317 (41.9%) |
| Diabetes | 316 | 757 | C0 (no): 649 (85.7%), C1 (yes): 108 (14.3%) |
| Hypertension | 124 | 288 | C0 (no): 196 (68.1%), C1 (yes): 92 (31.9%) |
| Nicotine | 316 | 757 | C0 (non-smoker): 407 (53.8%), C1 (smoker): 350 (46.2%) |

### 2.2 Motion Extraction (Optical Flow)

Motion can be extracted using either sparse feature tracking or dense optical flow. Sparse tracking detects and follows discrete keypoints, but this is unreliable in coronary angiography: vessels are thin and low-contrast, providing few stable features, and keypoints drift or disappear as contrast washes in and out across frames. Dense optical flow, in contrast, estimates pixel-wise displacement directly from temporal intensity variations, providing a more complete and coherent motion representation without relying on handcrafted keypoints. It is therefore better suited to these imaging conditions.

Following Yin et al. [10], we employ a Recurrent All-Pairs Field Transforms (RAFT) model [12], initialized from pretrained weights and fine-tuned on a synthetic coronary motion dataset (Section 3.1), to estimate flow between consecutive vessel-enhanced frames. Unlike prior angiography-specific approaches that rely on segmentation masks or selected frame pairs, our pipeline operates directly on long sequences without segmentation labels.

### 2.3 Motion Representation Analysis

We extract motion representations from dense optical-flow sequences using video-based deep neural networks. Stacked optical-flow fields from consecutive frames are fed into a motion representation network that captures spatial and temporal characteristics. Each 9-frame clip yields 8 flow fields, matching the temporal input length of the two-stream network.

We evaluate two architectures: a **two-stream network with a ResNet-18 backbone** [13] and a transformer-based **TimeSformer model** [14]. ResNet-18 serves as a lightweight convolutional baseline for optical-flow-based motion analysis, providing a strong spatial inductive bias and robustness for moderate-sized medical datasets. TimeSformer is included as a transformer-based video model

to assess whether self-attention can better capture global spatiotemporal motion dependencies. This comparison enables evaluation of convolutional versus attention-based representations for coronary motion analysis. The performance of these architectures in predicting clinical parameters from motion-derived representations is presented in Section 3.3.

During training, $224 \times 224$ patches are randomly cropped from the $512 \times 512$ optical-flow frames to reduce overfitting. The networks are trained to predict categorical clinical parameters from motion alone; these labels serve only for evaluation and association analysis, not as motion annotations. This lets us assess whether the learned motion representations encode clinically meaningful information without requiring manual motion annotations.

## 3 Results and Discussion

### 3.1 Synthetic Data Evaluation

Ground truth optical flow is rarely available in XCA, so we build a synthetic dataset with known motion for quantitative evaluation and fine-tuning.

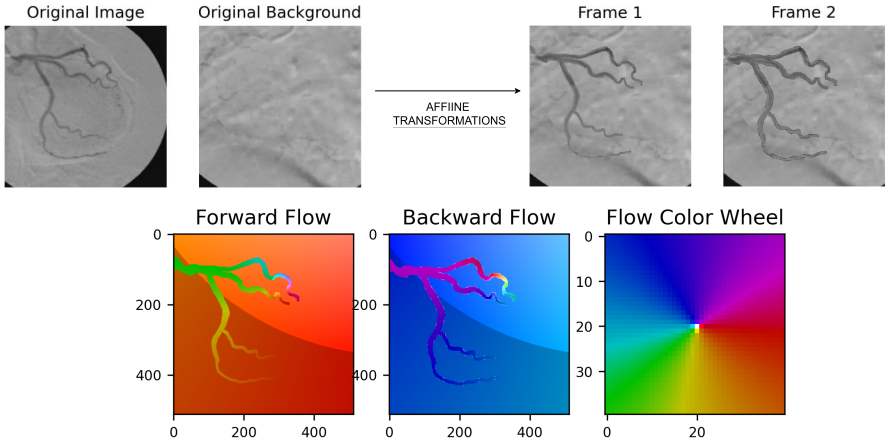


Fig. 1: Overview of the synthetic dataset generation process.

Synthetic sequences use segmented coronary artery images from [15], (134 vessel masks at $300 \times 300$), upscaled to $512 \times 512$ pixels to match the resolution of real angiographic data; randomly sampled empty frames from the real dataset serve as backgrounds to capture acquisition noise and minor camera motion. Motion is introduced via affine transformations including rotation, translation, scaling, and shear applied to vessel foregrounds and backgrounds.

Because all transformations are explicit, the forward and backward flow fields between consecutive frames are known exactly as ground-truth (Figure 1).

We evaluate qualitatively (Fig. 2) and quantitatively using endpoint error (EPE; Euclidean displacement error), flow loss (pixel-wise error), and pixel-threshold accuracy at 1, 3, and 5 pixels (percentage of pixels with EPE below the threshold). Under these controlled conditions, the estimated flow closely matches the ground truth: the model achieves a flow loss of 0.3238 px and an EPE of 0.5006 px, with 96.11%, 98.74%, and 99.23% of pixels within 1, 3, and 5 pixels, respectively.

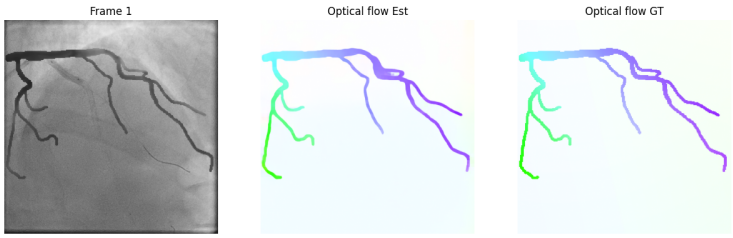


Fig. 2: The Optical Flow for the synthetic dataset, from left to right: synthetic frame 1, estimated optical flow, ground truth optical flow.

### 3.2 Real Data Motion Evaluation

Without ground-truth flow on real XCA, we assess quality through (i) *qualitative visualization* and (ii) *reconstruction-based metrics*. Both are performed on the analysis cohort screened with Frangi-based vessel enhancement and visual inspection (Sec. 2.1).

**Qualitative Evaluation (i)**: Fig. 3 shows a representative example of coronary motion extracted from a routine angiographic sequence. Panels A and B are two consecutive frames (t and t+1), between which dense optical flow is estimated following Frangi-based vessel enhancement. Panel C overlays the estimated motion, with color indicating the normalized magnitude of displacement along the vessel structure. The motion is concentrated along the coronary vessels rather than the background, consistent with the flow capturing vessel-aligned motion.

**Quantitative Evaluation via Reconstruction (ii)**: The estimated optical flow is used to warp the first frame and generate a reconstruction of the second frame. This reconstructed frame is then compared with the original second frame using Mean Squared Error (MSE), Intersection-over-Union (IoU), and sensitivity. As a baseline, we directly use the first frame as an approximation of the second frame, without applying optical flow. On real data, reconstruction reduces MSE from 1944.9 to 1710.6 and improves sensitivity from 0.49 to 0.55, with comparable IoU (0.34 versus 0.33). This trade-off is illustrated qualitatively in Fig. 4, where the reconstruction recovers a larger intersection (navy) region

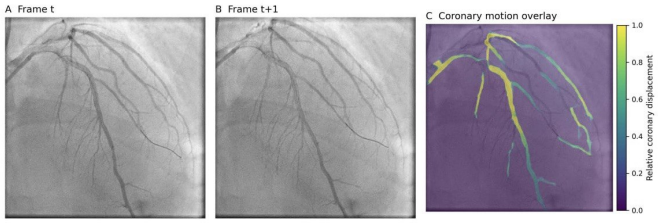


Fig. 3: Example of coronary artery motion extracted between two consecutive angiographic frames using the final fine-tuned model.

along the vessels. On synthetic data, short-horizon reconstruction between consecutive frames can be near-trivial, so we generate frame pairs with accumulated incremental affine motion; under this setting, reconstruction reduces MSE from 1933.9 to 1791.8 and increases sensitivity from 0.46 to 0.50, with a slight IoU drop from 0.29 to 0.27. The small IoU decrease stems from the warping itself: because the estimated flow is float-valued, a source point $(x, y)$ with flow $(u, v)$ lands at the non-integer location $(x + u, y + v)$ and is scattered to its four integer-pixel neighbors, $(\lfloor x+u \rfloor, \lfloor y+v \rfloor)$, $(\lfloor x+u \rfloor, \lceil y+v \rceil)$, $(\lceil x+u \rceil, \lfloor y+v \rfloor)$, and $(\lceil x+u \rceil, \lceil y+v \rceil)$, to avoid holes in the reconstruction. This scattering slightly enlarges the warped region and lowers IoU, even as MSE and sensitivity improve. Overall, reconstruction yields lower MSE and higher sensitivity than the baseline on both datasets, indicating that the flow captures meaningful motion; the comparable performance between the real and synthetic reconstructions further suggests that the optical-flow quality on real data is satisfactory.

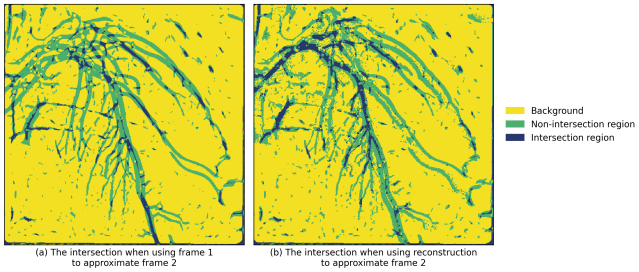


Fig. 4: Vessel overlap before (a) and after (b) optical-flow reconstruction, relative to the target frame. Yellow: background; green: non-intersection; navy: intersection. Greater overlap in (b) demonstrates improved alignment.

Table 2: Best training hyperparameter configuration for the motion representation networks.

| Optimizer | LR | Weight decay | Dropout | Patience | Loss |
|---|---|---|---|---|---|
| AdamW | $3 \times 10^{-5}$ | 0.1 | 0.5 | 20 ep. | Weighted BCE/CE* |

*BCE for binary parameters; cross-entropy for the three-class cholesterol task.

Table 3: Balanced accuracy (BA) and weighted F1-score for TimeSformer (depth=4) and two-stream ResNet-18, together with class ratios for each clinical parameter.

| | TimeSformer | | ResNet-18 | | |
|---|---|---|---|---|---|
| **Field** | **BA (%)** | **F1 (%)** | **BA (%)** | **F1 (%)** | **Class ratio** |
| Ejection Fraction | 78.31 | 78.18 | 82.83 | 84.31 | 66:34 |
| Cholesterol | 52.15 | 47.00 | 55.93 | 61.87 | 25:33:42 |
| Diabetes | 75.53 | 77.50 | 81.83 | 84.28 | 86:14 |
| Hypertension | 81.67 | 93.94 | 78.33 | 68.63 | 68:32 |
| Nicotine | 65.01 | 60.63 | 63.89 | 51.17 | 54:46 |

### 3.3 Motion Representation Analysis

In this section, we assess whether the learned motion representations encode clinically relevant information. Both architectures are trained with the configuration in Table 2. Because all five clinical parameters are class-imbalanced, we report balanced accuracy, which averages per-class recall and penalizes majority-class shortcuts, together with the F1-score.

As shown in Table 3, performance on all parameters exceeds chance level (50% for the binary parameters and 33% for the three-class cholesterol task). The Diabetes split, with a class distribution of 86:14, highlights why balanced accuracy is necessary: a model that always predicts the majority class would reach 86% accuracy without identifying any diabetic patients. Balanced accuracy avoids this by averaging performance across both classes. These results suggest the representations capture clinically relevant signal rather than class imbalance. ResNet-18 achieves higher balanced accuracy and F1-score on three of the five parameters (Ejection Fraction, Cholesterol, and Diabetes), while TimeSformer performs better on Hypertension and Nicotine.

## 4 Conclusion and Future work

This work demonstrates that optical flow can extract meaningful coronary motion patterns from routine angiographic sequences, supported by qualitative inspection and reconstruction-based evaluation on both synthetic and real data. Using learned optical flow–based motion representations, we observe measurable associations with several clinical parameters, including ejection fraction, cholesterol level, hypertension, and nicotine use. Although motion alone is insufficient

for standalone diagnosis, it carries clinically relevant information for cardiovascular risk analysis, establishing a scalable foundation for motion-based analysis of coronary angiography. Future work could combine these motion-derived representations with complementary anatomical biomarkers (e.g., stenosis severity) and established risk factors to build robust composite models.

**Prospects of Application:** This framework enables segmentation-free extraction of coronary motion from routine angiography, supporting large-scale retrospective studies of motion-derived functional signals. In research and clinical-assessment contexts, it may reveal associations with cardiovascular risk factors, aid patient stratification, and inform future cardiovascular risk-association tools without additional imaging or manual vessel annotation.

**Acknowledgments.** This project is supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 1011194785. The JU receives support from the European Union's Horizon Europe research and innovation programme and COCIR, EFPIA, Europa Bio, MedTech Europe, and Vaccines Europe.

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

## References

1. B. Chong, J. Jayabaskaran, S. M. Jauhari, S. P. Chan, R. Goh, M. Kueh, H. Li, Y. Chin, G. Kong, V. V. Anand, J. Wang, M. Muthiah, V. Jain, A. Mehta, S. L. Lim, R. Foo, G. A. Figtree, S. J. Nicholls, M. Mamas, J. Januzzi, N. W. S. Chew, A. M. Richards, and M. Chan, "Global burden of cardiovascular diseases: projections from 2025 to 2050." *European journal of preventive cardiology*, 2024.
2. K. Yamamoto, M. Natsuaki, T. Morimoto, H. Shiomi, Y. Takeji, K. Yamaji, Y. Matsumura-Nakano, Y. Yoshikawa, E. Yamamoto, M. Fuki, E. T. Kato, Y. Yamashita, H. Yaku, K. Nakatsuma, H. Watanabe, M. Matsuda, T. Tamura, M. Inoko, H. Mabuchi, M. Shirotani, H. Eizawa, K. Ishii, T. Inada, T. Onodera, E. Shinoda, T. Yamamoto, M. Toyofuku, M. Takahashi, H. Sakamoto, S. Miki, T. Aoyama, S. Suwa, Y. Sato, Y. Furukawa, K. Ando, K. Kadota, Y. Nakagawa, and T. Kimura, "Coronary artery disease without standard cardiovascular risk factors." *The American journal of cardiology*, 2021.
3. H. Gong, Z. Ahmed, S. Chang, E. K. Koons, J. Thorne, P. Rajiah, T. A. Foley, J. G. Fletcher, C. McCollough, and S. Leng, "Motion artifact correction in cardiac ct using cross-phase temporospatial information and synergistic attention gate and spatial transformer sub-networks," *Physics in Medicine  Biology*, vol. 69, 2024.
4. N. Hosadurg, K. Harrison, J. D. K. Nguyen, P. F. Rodriguez Lozano, C. M. Kramer, P. Norton, A. R. Patel, and T. C. Villines, "Impact of an institutional process change adopting end-systolic coronary cta acquisition and automated dose selection on patient throughput and image quality," *Journal of Cardiovascular Computed Tomography*, 2024.
5. E. Oubel, M. D. Craene, A. Hero, A. Pourmorteza, M. Huguet, G. Avegliano, B. Bijnens, and A. F. Frangi, "Cardiac motion estimation by joint alignment of tagged mri sequences," *Medical Image Analysis*, vol. 16, pp. 339–350, 2012.

6. M. Fei, M. Li, H. Ran, Z. Sheng, J. Dong, and P. Zhang, "Four-dimensional quantification on left atrial volume-strain in coronary heart disease patients without regional wall motion abnormalities: Correlation with the severity of coronary stenosis," *Echocardiography*, vol. 39, pp. 758–767, 2022.
7. J. Puentes, C. Roux, M. Garreau, and J. Coatrieux, "Dynamic feature extraction of coronary artery motion using dsa image sequences," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 857–871, 1998.
8. Z. Ding and M. Friedman, "Dynamics of human coronary arterial motion and its potential role in coronary atherogenesis," *Journal of Biomechanical Engineering*, vol. 122, pp. 488–492, 2000.
9. T. Konta and J. H. Bett, "Patterns of coronary artery movement and the development of coronary atherosclerosis," *Circulation Journal*, vol. 67, no. 10, pp. 846–850, 2003.
10. X.-L. Yin, D.-X. Liang, L. Wang, J. Xu, D. Han, K. Li, Z.-Y. Yang, J.-H. Xing, J.-Z. Dong, and Z.-Y. Ma, "Optical flow estimation of coronary angiography sequences based on semi-supervised learning," *Computers in Biology and Medicine*, vol. 146, p. 105663, 2022.
11. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, W. M. Wells, A. Colchester, and S. Delp, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 130–137.
12. Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II.* Berlin, Heidelberg: Springer-Verlag, 2020, p. 402–419. [Online]. Available: https://doi.org/10.1007/978-3-030-58536-5_24
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
14. G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Icml*, vol. 2, no. 3, 2021, p. 4.
15. F. Cervantes-Sanchez, I. Cruz-Aceves, A. Hernández-Aguirre, M. Hernández-González, and S. Solorio-Meza, "Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks," *Applied Sciences*, 2019.