Spatio-temporal Contrast Network for Data-efficient Learning of Coronary Artery Disease in Coronary CT Angiography

Xinghua Ma[1], Mingye Zou[1], Xinyan Fang[1], Yang Liu[1], Gongning Luo ✉[1,3], Wei Wang[2], Kuanquan Wang[1], Zhaowen Qiu[4], Xin Gao[3], and Shuo Li[5,6]

[1] the Faculty of Computing, Harbin Institute of Technology, Harbin, China
luogongning@hit.edu.cn
[2] the Faculty of Computing, Harbin Institute of Technology, Shenzhen, China
[3] the Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[4] the Institute of Information and Computer Engineering, Northeast Forestry University, Harbin, China
[5] the Department of Computer and Data Science, Case Western Reserve University, Cleveland, USA
[6] the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, USA

Abstract. Coronary artery disease (CAD) poses a significant challenge to cardiovascular patients worldwide, underscoring the crucial role of automated CAD diagnostic technology in clinical settings. Previous methods for diagnosing CAD using coronary artery CT angiography (CCTA) images have certain limitations in widespread replication and clinical application due to the high demand for annotated medical imaging data. In this work, we introduce the Spatio-temporal Contrast Network (SC-Net) for the first time, designed to tackle the challenges of data-efficient learning in CAD diagnosis based on CCTA. SC-Net utilizes data augmentation to facilitate clinical feature learning and leverages spatio-temporal prediction-contrast based on dual tasks to maximize the effectiveness of limited data, thus providing clinically reliable predictive results. Experimental findings from a dataset comprising 218 CCTA images from diverse patients demonstrate that SC-Net achieves outstanding performance in automated CAD diagnosis with a reduced number of training samples. The introduction of SC-Net presents a practical data-efficient learning strategy, thereby facilitating the implementation and application of automated CAD diagnosis across a broader spectrum of clinical scenarios. The source code is publicly available at the following link (https://github.com/PerceptionComputingLab/SC-Net).

Keywords: Coronary Artery Disease · Coronary CT Angiography · Data-efficient Learning · Spatio-temporal · Computer-aided Diagnosis.

1 Introduction

Automated diagnosis of coronary artery disease (CAD) using coronary CT angiography (CCTA) imaging is crucial in clinical settings for computer-aided diagnosis [5,7]. CAD has been widely recognized as one of the most serious medical conditions, frequently resulting in fatal outcomes such as heart attacks and heart failure [11]. Through CCTA, a non-invasive imaging technique, physicians can examine both the anatomical structure and blood flow in the coronary arteries [1]. Automating this diagnostic process offers physicians a quicker and more effective solution for accurately assessing CAD.

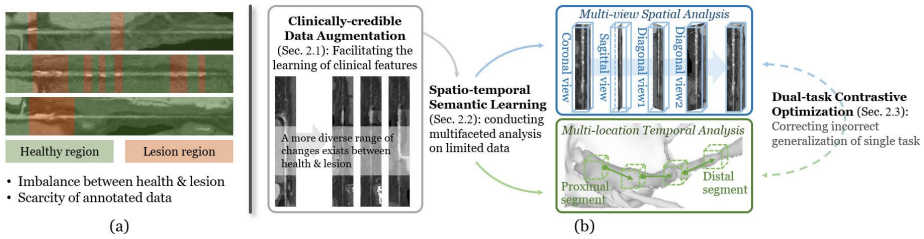

Fig. 1. (a) The scarcity of annotated data worsens the imbalance between healthy and lesion regions. (b) Spatio-temporal learning, along with dual-task contrast and data augmentation, maximize the use of limited samples, addressing data scarcity scenarios.

Deep learning-based computer-aided diagnostic technology has achieved clinical accuracy standards, but the high demand for annotated data for effective training has brought certain limitations to its adoption in clinical environments. Initially, CAD computer-aided diagnosis relied on semi-automated technologies [17, 20]. While these methods did not require annotated data for training, they were limited by the need for manual intervention from healthcare professionals, hindering their usability. Advancements in deep learning have led to the rise of automated CAD diagnosis [14], primarily focusing on sampling-point classification and object detection. Sampling-point classification involves identifying and classifying centerline points of coronary arteries within Curved Planar Reformation (CPR) volumes. Zreik *et al.* [19] pioneered the integration of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for feature extraction, enabling multi-class classification of stenoses and plaques. Subsequent researchers [3, 8] have made significant enhancements, especially in detecting significant stenoses with luminal narrowing $\leq 50\%$. Object detection, on the other hand, directly predicts regions of interest (ROIs) indicating lesions and their corresponding categories within CPR volumes. Zhang *et al.* [18] analyzed CPR imaging data and employed Faster R-CNN [4] for object detection task related to coronary lesions. Although these methods have achieved notable performance improvements, their replication and application in clinical settings are still hindered by the high demand for annotated data used for training.

The scarcity of annotated medical images poses a significant barrier to the extensive application of the aforementioned methods in clinical settings, while data-efficient learning also presents numerous challenges in the field of automated CAD diagnosis, as shown in Fig.1(a). (1) In clinical scenarios, there exists an imbalance between the proportion of coronary artery lesions and healthy regions [9]. The significance of lesions in automated diagnosis is further diminished by the limited volume of data and annotations, complicating the capture of variable lesions. (2) Insufficient training data significantly undermines the clinical reliability of previously data-intensive techniques. Within the confines of limited samples, single-task diagnostic methods encounter challenges in rectifying erroneous generalizations, thereby elevating the probability of missed and misdiagnosis. The challenges faced by data-efficient learning in automated CAD diagnosis collectively hinder its practical application and development in this clinical context.

In this work, we introduce the Spatio-temporal Contrast Network (SC-Net) for data-efficient learning of CAD diagnosis in CCTA. SC-Net aims to provide clinically reliable lesion assessment guidance despite limited data (Fig.1(b)), supported by three designs: (1) Clinically-credible data augmentation separates and recombines lesions within CPR volumes based on clinical guidance, promoting clinical learning and addressing sample imbalance in data scarcity scenarios. (2) Spatio-temporal semantic learning captures temporal and spatial relationships within CPR volumes, enabling comprehensive analysis with limited samples. (3) Dual-task contrastive optimization promotes knowledge sharing and corrects erroneous generalizations, reducing missed and misdiagnosis in data-efficient learning. Overall, our main contributions are summarized as follows: • SC-Net addresses, for the first time, the data-efficient learning challenge in automated CAD diagnosis, promoting the application of CAD computer-aided diagnosis in a wider range of clinical scenarios. • Spatio-temporal learning, incorporating dual-task contrast and clinically credible data augmentation, extracts comprehensive lesion-related features from limited data, improving CAD diagnosis reliability. • A comprehensive evaluation on a dataset of 218 cases demonstrates that SC-Net outperforms State-of-the-Art (SOTA) diagnostic methods, even with smaller training sets.

2 Method

SC-Net effectively utilizes limited data to accomplish data-efficient learning for automated CAD diagnosis in CCTA, as shown in Fig.2. It consists of three components: clinically-credible data augmentation, spatio-temporal semantic learning, and dual-task contrastive optimization.

2.1 Clinically-credible Data Augmentation

Clinically-credible data augmentation ensures that the training data better reflects the diversity and complexity of coronary lesions, thus effectively enabling

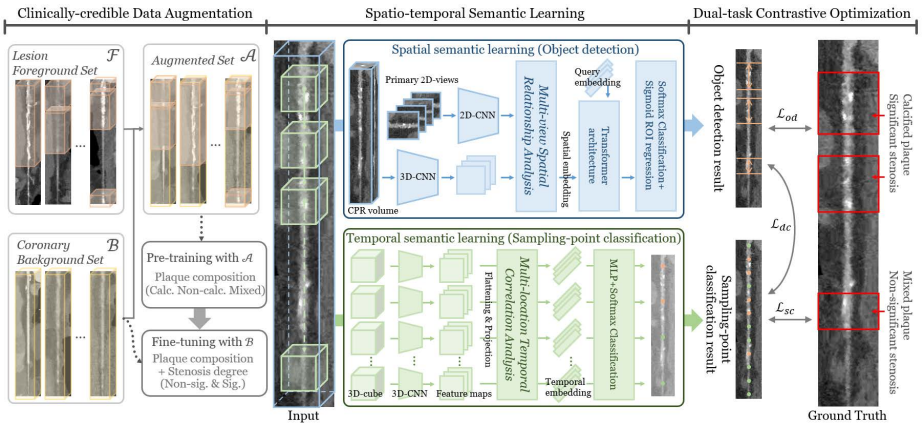

Fig. 2. SC-Net uses data augmentation to enhance the diversity of lesions with limited data, dual-task learning to extract spatio-temporal features, and prediction contrast to correct erroneous generalizations.

SC-Net to learn clinically valuable features for CAD diagnosis. According to guidance from the Society of Cardiovascular Computed Tomography (SCCT) [6], coronary artery stenosis leading to CAD results from the deposition of various components that form plaques attached to the blood vessel wall. Therefore, when using Computer Vision (CV) technology for automated CAD diagnosis, plaques with contrasting CT values are more distinguishable than subtle changes in stenosis. Based on these considerations, we augment the relatively limited dataset while preserving the clinical characteristics of lesions.

Specifically, two sets are created: the coronary background set and the lesion foreground set. The coronary background set $\mathcal{B}$ comprises all CPR volumes identified as background elements. The lesion foreground set $\mathcal{F}$ comprises all lesion ROIs in CPR volumes identified as foreground elements. Subsequently, randomly selected foreground elements from set $\mathcal{F}$ are overlaid onto randomly selected background elements from set $\mathcal{B}$ to form samples a, constituting the augmented set $\mathcal{A}$:

$$\mathcal{A} = \{a \mid a = (b - b_I) \cup f_I\} \qquad s.t. \quad b \sim \mathcal{B}, f \sim \mathcal{F} \tag{1}$$

where f and b respectively denote foreground elements randomly selected from set $\mathcal{F}$ and background elements randomly selected from set $\mathcal{B}$, and I denotes the indices of the foreground elements f in the CPR volume. This procedure enriches the variety of changing scenarios across different pathological and healthy segments, thereby providing SC-Net with more CPR volumes capable of learning clinically relevant features. It is worth noting that, due to variations in the diameters of different coronary artery segments, the training process based on augmented data focuses solely on evaluating plaques causing stenosis. For the overall training of SC-Net, we first pre-train on augmented data $\mathcal{A}$ with plaque

composition, and then fine-tune SC-Net using clinical data $\mathcal{B}$, which includes both plaque composition and stenosis degree.

2.2 Spatio-temporal Semantic Learning

Spatio-temporal semantic learning synchronously analyzes both the spatial relationships among multiple views of CPR volumes and the temporal correlation among multiple locations, fully utilizing multifaceted information from limited data for the high-quality feature extraction. Specifically, the tasks of object detection and sampling-point classification correspond to spatial and temporal semantic learning, respectively.

In spatial semantic learning, a CPR volume $x_{cpr} \in \mathbb{R}^{D\times N_1\times N_1}$ along with its four primary 2D-views $x^i_{vw} \in \mathbb{R}^{D\times N_1}(i \in [0,3])$ (*i.e.*, projections along the sagittal and coronal axes, as well as two diagonal axes [10]) serve as the input for object detection. x_{cpr} and x^i_{vw}s are processed by a 3D-CNN and a 2D-CNN to extract CPR features f_{cpr} and 2D-view features $f^i_{vw}(i \in [0,3])$, respectively. These features are then jointly fed into *Multi-view Spatial Relationship Analysis* to integrate location information among multiple views. During this step, four 3D zero-feature maps $\hat{f}^i_{vw} \in \mathbb{R}^{D\times N_1\times N_1}(i \in [0,3])$ are generated, with 2D-view features filling corresponding positions for each $\hat{f}^i_{vw}$. Weights ws are assigned to $\hat{f}_{vw}$s and f_{cpr}, and these maps aggregated based on the weights of each view:

$$f_{spa} = (1 - w_{cpr})\sum_{i=0}^{3} w^i_{vw}\hat{f}^i_{vw} + w_{cpr}f_{cpr} \in \mathbb{R}^{C\times L\times H\times W} \tag{2}$$

The aggregated feature maps f_{spa} are then transformed into spatial embeddings emd_{spa}s using a Fully Connected (FC) layer. emd_{spa}s are queried using the Transformer architecture [16] based on randomly initialized Q query embeddings $emb^i_{que} \in \mathbb{R}^{512}(i \in [0,Q])$ [2]. For each emb_{que}, two Multi-Layer Perceptrons (MLPs) with a Sigmoid regression head and a Softmax classification head respectively are then employed for ROI localization and lesion characterization.

In temporal semantic learning, a series of 3D-cubes $x^i_{cub} \in \mathbb{R}^{N_2\times N_2\times N_2}(i \in [0,L])$ serve as the input for sampling-point classification. Each cube is fed into a shallow 3D-CNN [8] to extract local features, which are then transformed into temporal embeddings $emb^i_{tem} \in \mathbb{R}^{512}(i \in [0,31])$ through flattening and projection. The *Multi-location Temporal Correlation Analysis* step utilizes a Transformer encoder to integrate temporal information between different locations. Finally, each temporal embedding is processed by an MLP with a Softmax classification head to perform lesion assessment for each sampling point.

2.3 Dual-task Contrastive Optimization

Dual-task Contrastive Optimization jointly optimizes two tasks focusing on different perspectives, enabling mutual supervision and error correction in generalization during data-efficient learning. SC-Net's training objective $\mathcal{L}_{overall}$ includes object detection loss $\mathcal{L}_{od}$ for spatial semantic learning, sampling-point

classification loss $\mathcal{L}_{sc}$ for temporal semantic learning, and dual-task contrastive loss $\mathcal{L}_{dc}$ for mutual supervision of multiple perspectives:

$$\mathcal{L}_{overall} = \mathcal{L}_{od} + \mathcal{L}_{sc} + \mathcal{L}_{dc} \tag{3}$$

The object detection loss $\mathcal{L}_{od}$ optimizes spatial semantics learning and sub-task prediction at two stages. At the first phase, the bipartite matching between the Ground truth set y_{od} and the prediction result set $\hat{y}_{od} = \{\hat{y}_{od}^i\}_{i=1}^{Q}$ is accomplished using the Hungarian algorithm [15]. The prediction target of this sub-task is represented as $(\hat{c}_i, \hat{r}_i)$, where c_i is the category label and $\hat{r}_i \in [0,1]^2$ is a vector $[\hat{e}_i, \hat{w}_i]$ that defines RoI center coordinate and the weight in the CPR volume. A permutation of Q elements $\sigma \in \mathfrak{S}Q$ is searched [2] for with the lowest cost to solve the bipartite matching:

$$\hat{\sigma} = \underset{\sigma \in \mathfrak{S}_Q}{\arg\min} \textstyle\sum_i^Q [-\mathbb{1}_{\{c_i \neq \varnothing\}} \hat{p}_{\sigma(i)}(c_i) + \mathbb{1}_{\{c_i \neq \varnothing\}} \mathcal{L}_{roi}(r_i, \hat{r}_{\sigma(i)})] \tag{4}$$

where $\hat{p}_{\sigma(i)}(c_i)$ is the probability of category c_i, $\varnothing$ denotes no-object category, and the RoI loss $\mathcal{L}_{roi}(r_i, \hat{r}_{\sigma(i)})$ is a linear combination of l_1 loss and the Intersection over Union (IoU) loss [13], with hyperparameters λ_{iou} and $\lambda_{L1} \in \mathbb{R}$ (*i.e.*, $\lambda_{iou}\mathcal{L}_{iou}(r_i, \hat{r}_{\sigma(i)}) + \lambda_{L1}\|r_i - \hat{r}_{\sigma(i)}\|_1$). At the second phase, $\mathcal{L}_{od}$ is formulated as a linear combination of the negative log-likelihood and the RoI loss, based on the matched pairs:

$$\mathcal{L}_{od}(y_{od}, \hat{y}_{od}) = \textstyle\sum_{i=1}^{Q} [-\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbb{1}_{\{c_i \neq \varnothing\}} \mathcal{L}_{roi}(r_i, \hat{r}_{\hat{\sigma}(i)})] \tag{5}$$

The sampling-point classification loss $\mathcal{L}_{sc}$ optimizes temporal semantics learning and sub-task prediction. The cross-entropy evaluation for each sampling-point between the Ground Truth set $y_{sc} = \{c_{sc}^i\}_{i=1}^{L}$ and the prediction result set $\hat{y}_{sc} = \{\hat{y}_{sc}^i\}_{i=1}^{L}$ is defined as:

$$\mathcal{L}_{sc}(y_{sc}, \hat{y}_{sc}) = -L^{-1} \textstyle\sum_{t=1}^{L} (\sum_j c_{sc}^i log(\hat{y}_{sc}^i)) \tag{6}$$

where c_{sc}^i and $\hat{y}_{sc}^i$ are the one-hot encoding and the predicted probability for the j-th class of the i-th sampling-point. The dual-task contrastive loss $\mathcal{L}_{dc}$ facilitates optimizing two sub-tasks by employing each other's prediction results as ground truth, thereby enhancing data utilization efficiency through mutual supervision:

$$\mathcal{L}_{dc} = \mathcal{L}_{od}(\mathcal{C}(\hat{y}_{sc}), \hat{y}_{od}) + \mathcal{L}_{sc}(\mathcal{C}^{-1}(\hat{y}_{od}), \hat{y}_{sc}) \tag{7}$$

where $\mathcal{C}(\cdot)$ represents the transformation of sampling-point classification results into RoIs with their respective categories, and $\mathcal{C}^{-1}(\cdot)$ represents the transformation of object detection results into the categories of sampling-points.

3 Experiment

3.1 Experimental configuration

Dataset. The clinical CCTA scans were obtained from 218 patients (age: 57.4 ± 6.2 years, 163 males) from 2019 to 2022, annotated by experienced physicians.

Table 1. Quantitative comparisons (50%/100% data volume) show the superior performance of SC-Net in data-efficient learning for CAD diagnosis. (↑: Stenosis; ↓: Plaque)

Method	ACC	Prec	Recall	F1	Spec
Texture CLS. [10]	0.805 / 0.879	0.851 / 0.906	0.847 / 0.909	0.849 / 0.907	0.678 / 0.792
3D-RCNN [19]	0.847 / 0.886	0.875 / 0.911	0.885 / 0.909	0.879 / 0.910	0.748 / 0.811
2D-RCNN+PT. [3]	0.837 / 0.867	0.871 / 0.893	0.876 / 0.898	0.873 / 0.895	0.724 / 0.784
TR-Net [8]	0.812 / 0.893	0.860 / 0.913	0.854 / 0.914	0.856 / 0.914	0.681 / 0.826
Coro. RCNN [18]	0.864 / 0.919	0.892 / 0.936	0.894 / 0.939	0.893 / 0.937	0.773 / 0.872
SC-Net (Ours)	0.914 / **0.928**	0.939 / **0.942**	0.939 / **0.946**	0.938 / **0.944**	0.861 / **0.879**
3D-RCNN [19]	0.796 / 0.859	0.862 / 0.904	0.862 / 0.906	0.862 / 0.905	0.571 / 0.692
Coro. RCNN [18]	0.844 / 0.887	0.897 / 0.924	0.895 / 0.923	0.896 / 0.923	0.663 / 0.752
SC-Net (Ours)	0.903 / **0.912**	0.936 / **0.941**	0.934 / **0.939**	0.935 / **0.940**	0.784 / **0.816**

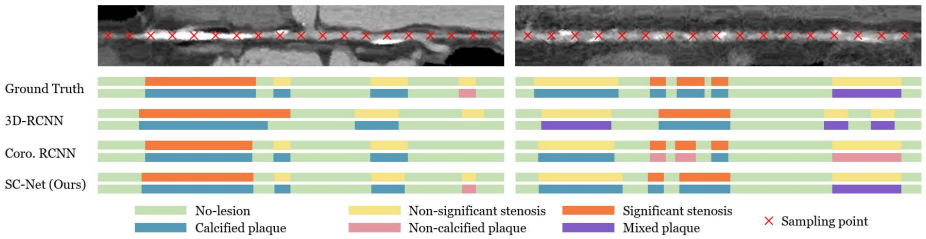

Fig. 3. Qualitative comparisons show the superiority of SC-Net in localizing and characterizing lesions within coronary segments.

A total of 1163 CPR volumes of the main coronary artery branches were reconstructed with the marching cube algorithm [12]. There were 994 coronary lesions in these CPR volumes, including 678 non-significant stenoses (208 calcified, 119 non-calcified, and 351 mixed plaques) and 316 significant stenoses (107 calcified, 94 non-calcified, and 115 mixed plaques).

Implementation details. The input shapes of CPR volume and 2D-views are initialized to $256 \times 64 \times 64$ and 256×64, respectively. Each CPR volume samples 32 3D-cubes uniformly with an interval of 8 voxels, where the shape of each 3D-cube is initialized as $25 \times 25 \times 25$. The number of query embeddings Q is set to 16, assuming the number of lesions in a coronary segment is always less than or equal to 16. The hyperparameters λ_{iou} and λ_{L1} of the IoU loss are set to 2 and 5 respectively.

Evaluation metrics. For the training process, 70% of CCTA scans are utilized, evenly distributed across lesion categories. The remaining 30% are divided between validation and test sets. Model weights from the best-performing validation run after 200 epochs are utilized for performance evaluation on the test set. Evaluation metrics comprise mean accuracy (ACC), precision (Prec), specificity (Spec), F1-score (F1), and specificity (Spec) pertaining to stenosis degree and plaque components at the artery-level.

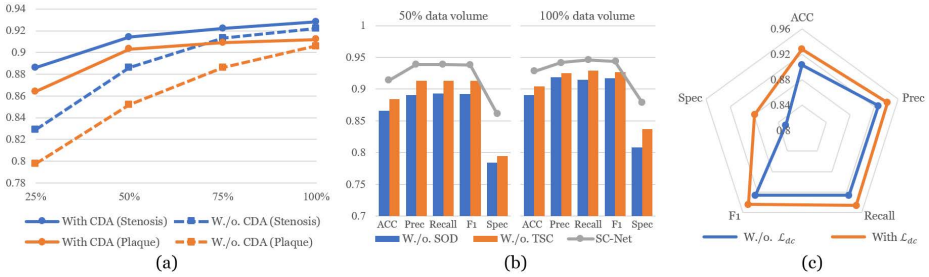

Fig. 4. Ablation analysis demonstrates the effectiveness of the SC-Net architecture for data-efficient learning.

3.2 Experimental results

Comparison with SOTAs. To evaluate the effectiveness of SC-Net in automated CAD diagnosis, we conducted a comparative experiment against representative SOTA methods cited in previous works [3, 8, 10, 18, 19]. Since plaques are not addressed in [3,8,10], we focused solely on evaluating the stenosis degree predicted by these methods during the evaluation process. As shown in Tab.1, the quantitative results demonstrate that SC-Net exhibits superior performance compared to SOTA methods, even when trained on a reduced dataset. SC-Net demonstrates significantly better performance compared to SOTA methods when using 100% of the training data. Even when the dataset is reduced to 50%, its performance remains comparable to those SOTA methods trained with 100% of the data. These findings suggest that our method enables more comprehensive analysis leveraging limited data resources, thereby enhancing the reliability of CAD diagnostic guidance in clinical settings. Moreover, qualitative results reveals that SC-Net consistently delivers clinically reliable and realistic automated CAD diagnostic outcomes, as shown in Fig.3. It effectively addresses diverse scenarios, including extensive stenosis caused by continuous plaques and various stenoses induced by multiple small plaques. Notably, there exists a strong correlation between the prediction of stenosis and plaques, leading to a reduction in the probability of physiologically implausible predictions [19].

Ablation study. To validate the effectiveness of SC-Net's architectural design, we conducted ablation analyses on three designs outlined in Sec. 2.1 to 2.3. The absence of Clinically-credible Data Augmentation (CDA) resulted in a significant decline in SC-Net's performance as the training set size decreased (Fig.4(a)), underscoring the augmentation's role in enhancing data diversity for learning clinical features. The differences in performance observed between the removal of spatially-related object detection (SOD) and temporally-related sampling-point classification (TSC) (Fig.4(b)) suggest that spatio-temporal semantic learning can effectively leverage limited data from various perspectives. Comparing SC-Net's performance with or without $\mathcal{L}_{dc}$ of dual-task contrastive optimization (Fig.4(c)), optimizing through prediction contrast of results from

the two tasks can better rectify generalization errors from limited samples, thereby further enhancing clinical reliability.

4 Conclusion

In this work, we proposed a novel data-efficient learning solution (SC-Net), which addresses the challenge of data scarcity in automated CAD diagnosis based on CCTA for the first time. Compared to previous technologies focusing on sampling-point classification and object detection, SC-Net leverages dual tasks to extract spatiotemporal semantics, thereby mutually enhancing prediction through contrast. Our experimental results demonstrated that SC-Net outperforms SOTA methods in data-efficient learning, achieving superior performance with less training data. Through clinically credible data amplification and multifaceted analysis, SC-Net offers a cost-effective and reliable CAD diagnostic solution across a broader range of clinical scenarios.

Acknowledgments. This work was supported by the Key Research & Development Program of Heilongjiang Province under Grant 2023X01A08, the National Natural Science Foundation of China under Grants 62272135, 62372135, and the King Abdullah University of Science and Technology (KAUST) Office of Research Administration (ORA) under Awards No. FCC/1/1976-44-01, FCC/1/1976-45-01, and REI/1/5234-01-01.

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

References

1. Abdelrahman, K.M., Chen, M.Y., Dey, A.K., Virmani, R., Finn, A.V., Khamis, R.Y., Choi, A.D., Min, J.K., Williams, M.C., Buckler, A.J., et al.: Coronary computed tomography angiography from clinical uses to emerging technologies: Jacc state-of-the-art review. Journal of the American College of Cardiology **76**(10), 1226–1243 (2020)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. pp. 213–229. Springer (2020)
3. Denzinger, F., Wels, M., Ravikumar, N., Breininger, K., Reidelshöfer, A., Eckert, J., Sühling, M., Schmermund, A., Maier, A.: Coronary artery plaque characterization from ccta scans using deep learning and radiomics. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. pp. 593–601. Springer (2019)

4. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
5. Kagiyama, N., Shrestha, S., Farjo, P.D., Sengupta, P.P.: Artificial intelligence: practical primer for clinical research in cardiovascular disease. Journal of the American Heart Association **8**(17), e012788 (2019)
6. Leipsic, J., Abbara, S., Achenbach, S., Cury, R., Earls, J.P., Mancini, G.J., Nieman, K., Pontone, G., Raff, G.L.: Scct guidelines for the interpretation and reporting of coronary ct angiography: a report of the society of cardiovascular computed tomography guidelines committee. Journal of cardiovascular computed tomography **8**(5), 342–358 (2014)
7. Luo, G., Ma, X., Guo, J., Zou, M., Wang, W., Cao, Y., Wang, K., Li, S.: Trajectory-aware adaptive imaging clue analysis for guidewire artifact removal in intravascular optical coherence tomography. IEEE Journal of Biomedical and Health Informatics (2023)
8. Ma, X., Luo, G., Wang, W., Wang, K.: Transformer network for significant stenosis detection in ccta of coronary arteries. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 516–525. Springer (2021)
9. North, B.J., Sinclair, D.A.: The intersection between aging and cardiovascular disease. Circulation research **110**(8), 1097–1108 (2012)
10. Tejero-de Pablos, A., Huang, K., Yamane, H., Kurose, Y., Mukuta, Y., Iho, J., Tokunaga, Y., Horie, M., Nishizawa, K., Hayashi, Y., et al.: Texture-based classification of significant stenosis in ccta multi-view images of coronary arteries. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 732–740. Springer (2019)
11. Pagliaro, B.R., Cannata, F., Stefanini, G.G., Bolognese, L.: Myocardial ischemia and coronary disease in heart failure. Heart Failure Reviews **25**(1), 53–65 (2020)
12. Rajon, D., Bolch, W.E.: Marching cube algorithm: review and trilinear interpolation adaptation for image-based dosimetric models. Computerized Medical Imaging and Graphics **27**(5), 411–435 (2003)
13. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 658–666 (2019)
14. Singh, S.P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., Gulyás, B.: 3d deep learning on medical images: a review. Sensors **20**(18), 5097 (2020)
15. Stewart, R., Andriluka, M., Ng, A.Y.: End-to-end people detection in crowded scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2325–2333 (2016)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems. pp. 5998–6008 (2017)
17. Xu, Y., Liang, G., Hu, G., Yang, Y., Geng, J., Saha, P.K.: Quantification of coronary arterial stenoses in cta using fuzzy distance transform. Computerized Medical Imaging and Graphics **36**(1), 11–24 (2012)
18. Zhang, Y., Ma, J., Li, J.: Coronary r-cnn: Vessel-wise method for coronary artery lesion detection and analysis in coronary ct angiography. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Con-

ference, Singapore, September 18–22, 2022, Proceedings, Part III. pp. 207–216. Springer (2022)
19. Zreik, M., Van Hamersvelt, R.W., Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography. IEEE transactions on medical imaging **38**(7), 1588–1598 (2018)
20. Zuluaga, M.A., Magnin, I.E., Hernández Hoyos, M., Delgado Leyton, E.J., Lozano, F., Orkisz, M.: Automatic detection of abnormal vascular cross-sections based on density level detection and support vector machines. International journal of computer assisted radiology and surgery **6**, 163–174 (2011)