Clinical Risk-Aware Multi-Level Grading for Coronary Artery Stenosis through Curved Feature Reconstruction

Shishuang Zhao[*1] 🆔   SHISHUANG.ZHAO@YIZHUN-AI.COM
Hongtai Li[*2]   LIHONGTAI999@163.COM
Junjie Hou[1]   JUNJIE.HOU@YIZHUN-AI.COM
Yuhang Liu[†1] 🆔   YUHANG.LIU@YIZHUN-AI.COM

[1] *Yizhun Medical AI Co., Ltd*

[2] *The First Hospital of China Medical University*

Abstract

Developing a multi-level grading model for coronary artery stenosis holds great clinical significance for the diagnosis of coronary artery disease. However, designing an effective multi-level deep learning algorithm faces significant challenges. Specifically, utilizing CCTA or 3D SCPR images alone presents inherent shortcomings: CCTA images are difficult to analyze due to the tortuous paths of blood vessels, while 3D SCPR images are prone to abnormal distortions that hinder accurate grading. Furthermore, different stenosis grades are associated with varying clinical risks, and incorporating this association into the algorithm is non-trivial. To address the former problems, we propose the Curved Feature Reconstruction (CFR) module, which uses vessel curves as prior and employs a point-by-point correspondence strategy to precisely align and fuse features from both 3D SCPR and CCTA images. Meanwhile, a Clinical Risk-Aware (CR) Loss is employed to introduce clinical risk relevance into the network training so that the algorithm can better align with the clinical diagnosis. The experimental results on a in-house dataset reveal that our approach significantly outperforms other methods, and several ablation studies also demonstrate the effectiveness of our proposed designs.

Keywords: Multi-level stenosis grading, Deep learning, CCTA, 3D SCPR

1. Introduction

Coronary artery stenosis, which refers to the narrowing of coronary arteries, has been recognized as the primary indicator of coronary artery disease (CAD)(Jensen et al., 2020; Naghavi et al., 2003; Otsuka et al., 2013). Accurate grading of stenosis is crucial for the effective diagnosis of CAD through Coronary Computed Tomographic Angiography (CCTA), a widely used non-invasive diagnostic examination(Mowatt et al., 2008). In clinical practice, stenosis can be graded on a 5-level scale, from 1 to 5, representing the increasing degrees of stenosis: minimal (1–24%), mild (25–49%), moderate (50–69%), severe stenosis (70–99%), and occluded vessel (100%)(Cury et al., 2022). Different stenosis grades correspond to different heart health risks, which leads to various medical interventions and treatments. Therefore, achieving accurate multi-level stenosis grading is a significant step towards the development of automatic computer-aided diagnosis system for CAD.

---

[*] Contributed equally
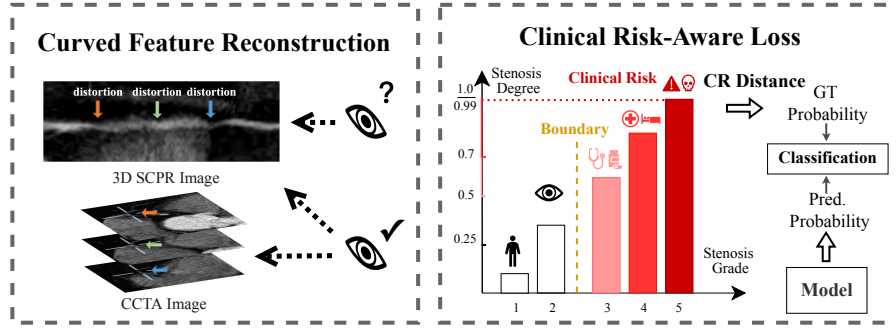
[†] Corresponding Author

Figure 1: Illustrations of the challenges and our solutions. In the left figure, identifying suspicious stenosis solely through 3D SCPR images is difficult due to abnormal distortions (indicated by colored arrows), while inclusion of CCTA images can provide more comprehensive information for stenosis grading. In the right figure, we propose CR Loss to encode the important clinical risk boundary explicitly into the model training.

To date, several studies have focused on multi-level stenosis grading(Zreik et al., 2018; Ma et al., 2021; Tejero-de Pablos et al., 2019; Kirişli et al., 2013; Kelm et al., 2011; Zhang et al., 2022). However, a common issue with these works is the insufficient incorporation of prior clinical knowledge, as shown in Figure 1. **Firstly**, clinicians primarily analyze CCTA images to diagnose CAD, while previous studies(Zreik et al., 2018; Ma et al., 2021; Zhang et al., 2022) have relied on 3D Straightened Curved Planar Reformation (3D SCPR) images, which may impede the analysis process due to abnormal distortions caused by the reformation mechanism, although 3D SCPR images provide information on the 3D geometry of arteries. **Secondly**, clinicians have different medical treatments for different stenosis grades, and the observation of moderate or greater stenosis serves as a critical diagnostic criterion for CAD(Cury et al., 2022). Patients with stenosis grade greater than 2 are at higher risk and may require further assessments and anti-anginal therapy, while those with grade 1 or 2 only need preventive pharmacotherapy(Cury et al., 2022). Thus, a clinical risk boundary exists between grade 1/2 and grade 3/4/5, and misclassification on either side of the boundary can have serious consequences. However, previous studies did not consider the clinical risk relevance of stenosis grading. They either focused on crude binary classification(Zreik et al., 2018; Ma et al., 2021; Tejero-de Pablos et al., 2019) or stenosis rate regression(Kirişli et al., 2013; Kelm et al., 2011; Zhang et al., 2022), treating different stenosis grades equally, which does not meet the demands of clinical practice. Additionally, accurately defining and acquiring stenosis rates is challenging due to the complex nature of the coronary lumen space.

To address the shortcomings of previous methods, we propose a novel deep learning-based approach for achieving multi-level stenosis grading. Our approach consists of two key designs to tackle the aforementioned problems. Firstly, we introduce the *Curved Feature Reconstruction* (CFR) module to address the limitations of 3D SCPR images. The CFR module integrates features from CCTA images with features from 3D SCPR images, using

vessel curve priors to establish a connection between the two image types. By combining the undistorted CCTA images, which retain the original texture details around stenosis regions, with 3D SCPR images, a more comprehensive analysis of stenosis can be achieved. Secondly, to align network training with clinical practice, we propose a novel loss function called *Clinical Risk-Aware Loss* (CR Loss). Specifically, we define the CR Distance to quantify the dissimilarity between different grades. We treat stenosis grading as a classification task, with the CR Distance serving as the metric function for the encoding of the ground truth probability distribution, explicitly incorporating clinical relevance into the network.

Our contributions are three-fold. (1) We propose the CFR module, which enables comprehensive stenosis analysis from both 3D SCPR and CCTA images. (2) We introduce the CR Loss, a novel loss function that explicitly encodes the important clinical risk boundary into network training. (3) We introduce two new metrics, MCRE and AMCRE, which measure the clinical risk relevance of stenosis grading results. Our method significantly outperforms other approaches in these metrics, indicating its better alignment with clinical needs.

2. Related Works

Regarding model input, most previous studies solely use 3D SCPR images as input(Ma et al., 2025, 2024, 2021; Zreik et al., 2018; Zhang et al., 2022; Candemir et al., 2020), which are prone to abnormal distortions and lacks the original texture details. In terms of methodology, previous methods for stenosis grading can be divided into two categories: stenosis rate regression(Kirişli et al., 2013; Kelm et al., 2011; Shahzad et al., 2013; Zhang et al., 2022) and binary classification(Zreik et al., 2018; Ma et al., 2021; Denzinger et al., 2019; Tejero-de Pablos et al., 2019). However, estimating vessel lumen accurately for stenosis rate calculation is a challenging task due to the complex nature of coronary lumen space, and the stenosis grade cannot be determined if the reference lumen is either non-existent or unreliable(Cury et al., 2022). Moreover, crude binary classification is not sufficient to meet the clinical requirements. Therefore, none of these approaches fulfills the requirements of the multi-level grading mechanism in clinical diagnosis. In this paper, we present the first deep learning-based multi-level stenosis grading method to facilitate the practical utilization.

3. Method

In this section, we will elaborate on the details of our method. Let $\mathbf{I} \in \mathbb{R}^{H \times W \times D}$, $\mathcal{S} = \{\mathbf{p}_i\}_{i=1}^{N}$ and $\overline{\mathcal{S}} = \{\mathbf{p}_i | \mathbf{p}_i \in \mathcal{S}\}_{i=s}^{e}$ denote the CCTA image, the centerline points, and stenosis instance respectively, where $\mathbf{p}_i \in \mathbb{R}^3$ is the 3D point location; N is the number of centerline points; and $1 \leq s < e \leq N$. The stenosis instance is a vessel segment consisting of several consecutive centerline points actually. Given a CCTA image and a stenosis instance $\overline{\mathcal{S}}$, the goal of the stenosis grading task is to predict the grade $y \in \{1, 2, 3, 4, 5\}$ for $\overline{\mathcal{S}}$.

Our system pipeline (Figure 2) can be divided into two parts: 1) the *Curved Feature Reconstruction* module to extract features from 3D SCPR image and CCTA patch and fuse them to predict the stenosis grade (Section 3.1); and 2) a novel *Clinical Risk-Aware Loss* to incorporate clinical relevance into the network training (Section 3.2). The following sections provide details of our method.
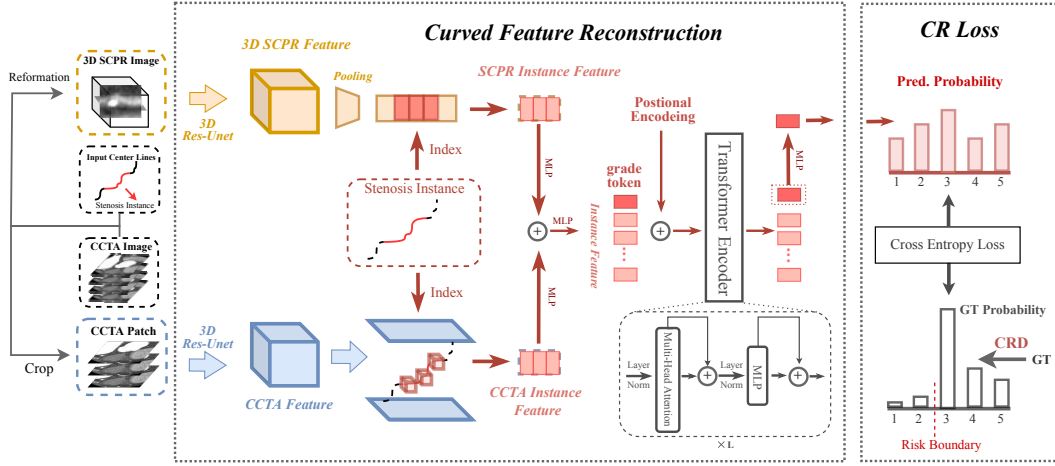
Figure 2: The overall framework of our method. Firstly, We obtain CCTA patch and 3D SCPR image as model inputs using CCTA image and input centerlines. Then CFR module is adopted to extract and fuse features from the inputs for stenosis grading. To incorporate clinical risk boundaries into neural network training, we finally apply a novel CR Loss for loss calculation.

3.1. Curved Feature Reconstruction

Previous works(Ma et al., 2021; Zreik et al., 2018; Zhang et al., 2022; Candemir et al., 2020) usually overlook the presence of abnormal distortions in 3D SCPR images, which may adversely affect stenosis grading. In this work, we incorporate the undistorted CCTA images into the network inputs to provide original texture around stenosis regions that may be lost in 3D SCPR images.

Firstly, as the stenosis grade is determined by comparing the stenotic part with the normal part, we extend stenosis instance $\overline{\mathcal{S}} = \{\mathbf{p}_i | \mathbf{p}_i \in \mathcal{S}\}_{i=s}^{e}$ along the vessel centerline to get $\hat{\mathcal{S}} = \{\mathbf{p}_i | \mathbf{p}_i \in \mathcal{S}\}_{i=s-m}^{e+m}$, where m is a hyper-parameter, to increase the vascular perceptual field. Then the 3D SCPR image $\mathbf{I}_{\mathrm{scpr}} \in \mathbb{R}^{H_{\mathrm{s}} \times W_{\mathrm{s}} \times D_{\mathrm{s}}}$ is obtained given the centerline points $\hat{\mathcal{S}}$ and the original CCTA image $\mathbf{I}$. Besides, we crop a patch $\mathbf{I}_{\mathrm{ccta}} \in \mathbb{R}^{H_{\mathrm{c}} \times W_{\mathrm{c}} \times D_{\mathrm{c}}}$ from CCTA image $\mathbf{I}$, which fully contains the vascular regions of $\hat{\mathcal{S}}$, to preserve more texture details. Finally, we feed $\mathbf{I}_{\mathrm{scpr}}$ and $\mathbf{I}_{\mathrm{ccta}}$ into the network to extract features separately and fuse them together to predict the stenosis grade.

Feature Extraction and Fusion. We utilize two separate 3D ResUNet(Yu et al., 2017) as backbones to extract 3D SCPR features $\mathbf{F}_{\mathrm{scpr}} \in \mathbb{R}^{H_{\mathrm{s}} \times W_{\mathrm{s}} \times D_{\mathrm{s}} \times C}$ and CCTA features $\mathbf{F}_{\mathrm{ccta}} \in \mathbb{R}^{H_{\mathrm{c}} \times W_{\mathrm{c}} \times D_{\mathrm{c}} \times C}$, where C is the channel size.

As $\mathbf{F}_{\mathrm{ccta}}$ has the same shape as $\mathbf{I}_{\mathrm{ccta}}$, the CCTA stenosis instance features can be directly acquired by indexing the stenosis instance coordinates $\mathbf{p}_i \in \overline{\mathcal{S}}$ from $\mathbf{F}_{ccta}$. We then use an MLP to project the instance features into new space and obtain $\overline{\mathbf{F}}_{\mathrm{ccta}} \in \mathbb{R}^{(e-s+1) \times C}$.

$$\overline{\mathbf{F}}_{\mathrm{ccta}} = \mathrm{MLP}\left(\left[\mathbf{F}_{\mathrm{ccta}}[\mathbf{p}_s], ..., \mathbf{F}_{\mathrm{ccta}}[\mathbf{p}_e]\right]\right) \tag{1}$$

4

where $\mathbf{p}_i \in \overline{\mathcal{S}}, \mathbf{F}_{\text{ccta}}[\mathbf{p}_i]$ means the local feature of $\mathbf{F}_{\text{ccta}}$ at position $\mathbf{p}_i$.

To obtain the 3D SCPR stenosis instance features, we first apply a 2D average pooling on $\mathbf{F}_{\text{scpr}}$ along the dimensions of H_s and W_s to get $\mathbf{F}'_{\text{scpr}} \in \mathbb{R}^{D_s \times C}$. Subsequently, the stenosis instance features are indexed from $\mathbf{F}'_{\text{scpr}}$ using the position of stenosis instance $\overline{\mathcal{S}}$ in points set $\hat{\mathcal{S}}$. And an MLP is then employed to generate $\overline{\mathbf{F}}_{\text{scpr}} \in \mathbb{R}^{(e-s+1) \times C}$.

$$\overline{\mathbf{F}}_{\text{scpr}} = \text{MLP}\left(\left[\mathbf{F}'_{\text{scpr}}[m], ..., \mathbf{F}'_{\text{scpr}}[m + e - s]\right]\right) \tag{2}$$

We then apply the element-wise addition on $\overline{\mathbf{F}}_{\text{ccta}}$ and $\overline{\mathbf{F}}_{\text{scpr}}$ to get the texture-enhanced stenosis instance features $\mathbf{F} = \overline{\mathbf{F}}_{\text{ccta}} + \overline{\mathbf{F}}_{\text{scpr}}$.

Grade Prediction. Transformer(Vaswani et al., 2017) has been proven to perform well in many computer vision tasks(Dosovitskiy et al., 2020; Liu et al., 2021). Here we also apply an attention-based network for stenosis grade prediction. Following (Dosovitskiy et al., 2020), we first add a learnable grade token $\mathbf{f}_{\text{grade}}$ to the stenosis instance features $\mathbf{F} = [\mathbf{f}^1, ..., \mathbf{f}^{e-s+1}]$, where $\mathbf{f}^i \in \mathbb{R}^C$ is the i-th token of $\mathbf{F}$, to form a sequence of tokens $[\mathbf{f}_{\text{grade}}, \mathbf{f}^1, ..., \mathbf{f}^{e-s+1}]$. Then the tokens are added by learnable position embeddings $\mathbf{E}_{pos} \in \mathbb{R}^{(e-s+2) \times C}$ and the resulting sequence serves as the input of the Transformer encoder, which contains multiple layers of multi-head self-attention (MSA) and MLP blocks.

$$\begin{aligned} \mathbf{g}_0 &= [\mathbf{f}_{\text{grade}}, \mathbf{f}^1, ..., \mathbf{f}^{e-s+1}] + \mathbf{E}_{pos} \\ \mathbf{g}'_l &= \text{MSA}(\text{LN}(\mathbf{g}_{l-1})) + \mathbf{g}_{l-1}, \ l = 1 ... L \\ \mathbf{g}_l &= \text{MLP}(\text{LN}(\mathbf{g}'_l)) + \mathbf{g}'_l, \ l = 1 ... L \\ \mathbf{g} &= \text{LN}(\mathbf{g}_L^0) \end{aligned} \tag{3}$$

where L is the number of layers in the encoder. The resulting feature $\mathbf{g} \in \mathbb{R}^C$ is then fed into an MLP to predict the final grade probability $\hat{P}$.

$$\hat{P} = \text{Softmax}(\text{MLP}(\mathbf{g})) \in \mathbb{R}^5 \tag{4}$$

3.2. Clinical Risk-Aware Loss

Clinically, patients with different stenosis grades have varying clinical risks, requiring various treatment methods(Cury et al., 2022). Therefore, an effective stenosis grading metric must be clinically risk-aware, applying a large penalty when predicted risk significantly deviates from the ground truth risk.

CR Distance. Following the above intuition, we propose a novel Clinical Risk-Aware Distance (CR Distance) as the distance metric to measure the dissimilarity between prediction and ground truth. CR Distance takes into account the clinical risks associated with stenosis grades, which can be formulated as follows:

$$\text{CRD}(a, b) = \begin{cases} (|a - b| + 1)^2, & a \le 2 < b \ \text{or} \ b \le 2 < a \\ |a - b|, & \text{else} \end{cases} \tag{5}$$

where a clinical risk boundary is between 1/2 and 3/4/5. When both inputs are on one side of the boundary, the CR Distance is equal to the average error as this type of mistake is not as significant. However, when inputs are on either side of the boundary, the CR Distance

takes the form of a square error, with a penalty term 1 added. As a result, the distance between the grades on either side of the boundary measured by CR Distance is much larger, which is compatible with the clinical characteristic of stenosis grade.

CR Loss. Stenosis grading is an ordinal regression problem and can be handled naturally by K-rank ordinal regression. However, it is hard to leverage the clinical risk boundary for K-rank ordinal regression as it actually performs binary classification during training. In our work, we approach stenosis grading as a classification task with a soft ground truth target(Diaz and Marathe, 2019) $P = [P_1, ..., P_5]$:

$$P_i = \frac{\exp(-\phi\,(gt, i))}{\sum_{j=1}^{5} \exp(-\phi\,(gt, j))}, \tag{6}$$

where $\phi(gt, i)$ is a metric function that measures the distance between the ground truth label gt and grade $i \in \{1, 2, 3, 4, 5\}$.

In this work, we propose to use the CR Distance as the metric function ϕ to introduce the clinical characteristics of stenosis into model training. As a result, the soft ground truth target can better reflect the clinically unacceptable confusion between stenosis grades 1/2 and 3/4/5. Our CR Loss is calculated by:

$$\text{CR Loss} = -w_{gt} \sum_{i=1}^{5} P_i \log(\hat{P}_i) = -w_{gt} \sum_{i=1}^{5} \frac{\exp(-\text{CRD}\,(gt, i))}{\sum_{j=1}^{5} \exp(-\text{CRD}\,(gt, j))} \log(\hat{P}_i) \tag{7}$$

where $w_{gt} \in \mathbf{w}$ is a loss weight for gt grade and $\mathbf{w} = [w_1, w_2, w_3, w_4, w_5]$ is a loss weights vector for all grades, as the number of each grade level is imbalanced. We also compare CR Distance with other metric functions (like absolute distance (AD) and square distance (SD)) in Table 3 to demonstrate its superiority.

4. Experiment

4.1. Experimental Settings

4.1.1. Dataset.

We collect a private dataset comprised of 500 CCTA scans, which contains labels of center-line points of coronary arteries and stenosis instances with corresponding grades annotated by at least three experienced radiologists. The dataset is collected in compliance with the terms of the licensing agreement and ethical certification. The number of stenosis instances in each grade level from 1 to 5 is 1654, 1443, 511, 266, and 40 respectively. We randomly split the dataset into train, validation and test set with a ratio of 3:1:1.

Evaluation metrics. Stenosis grading is essentially an ordinal regression task, and traditional evaluation metrics such as Quadratic Weighted Kappa (**QWK**)(Warrens, 2012), Mean Absolute Error (**MAE**), and Average Mean Absolute Error (**AMAE**) are naturally suited for it. However, these metrics do not take into account the clinical risk associated with each stenosis grade. CR distance incorporated with clinical risk boundary is suitable for measuring the clinical risk relevance of stenosis grading results. Therefore, we introduce two new metrics based on CRD: Mean Clinical Risk-Aware Error (**MCRE**) and Average Clinical Risk-Aware Error (**AMCRE**), MCRE is defined as the average of CRD over all

Table 1: Comparison with other methods.

Method	Data Type	QWK↑	MAE↓	AMAE↓	MCRE↓	AMCRE↓
Random Forest Regression	3D SCPR	0.378	0.686	1.141	2.167	4.555
Coronary R-CNN	3D SCPR	0.497	0.564	0.829	1.743	2.791
Multi-class Classification	3D SCPR	0.611	0.526	0.744	1.436	2.156
K-rank Ordinal Regression	3D SCPR	0.672	0.500	0.569	1.295	1.425
Ours	CCTA+3D SCPR	**0.712**	**0.438**	**0.533**	**1.185**	**1.186**

instances: $\mathrm{MCRE} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{CRD}(pred_i, gt_i)$ and AMCRE is calculated as the average of MCRE for all classes $\mathrm{AMCRE} = \frac{1}{C}\sum_{i=1}^{C}\mathrm{MCRE}_i$.

Implementation details. The loss weights $\mathbf{w}$, extend length m, channel size C, number of transformer layers L and attention head number are set to $[1, 1, 2, 3, 4]$, 10, 64, 8 and 4 by default. The proposed model is trained for 150 epochs with 15 epochs warm up and cosine learning rate decay schedule. AdamW(Loshchilov and Hutter, 2017) optimizer is adopted with learning rate of 0.001, weight decay of 0.001 and batch size of 32. All experiments are conducted on 8 NVIDIA GeForce RTX 3090 GPUs.

4.2. Experiment Results

Quantitative results. In Table 1, we compare our method with other approaches, including Random Forest Regression(Kelm et al., 2011) and Coronary R-CNN(Zhang et al., 2022). Several baselines based on hard label multi-class classification and K-rank ordinal regression are also provided for comparison. The results show that our approach outperforms all other methods by a large margin. Particularly, our method exhibits superior performance in the metrics of MCRE and AMCRE, which are relevant to distinguish stenosis grades between 1/2 and 3/4/5. This aligns with the design objective of CR loss indicating the clinical risk-aware ability has been encoded into the network through CR loss.
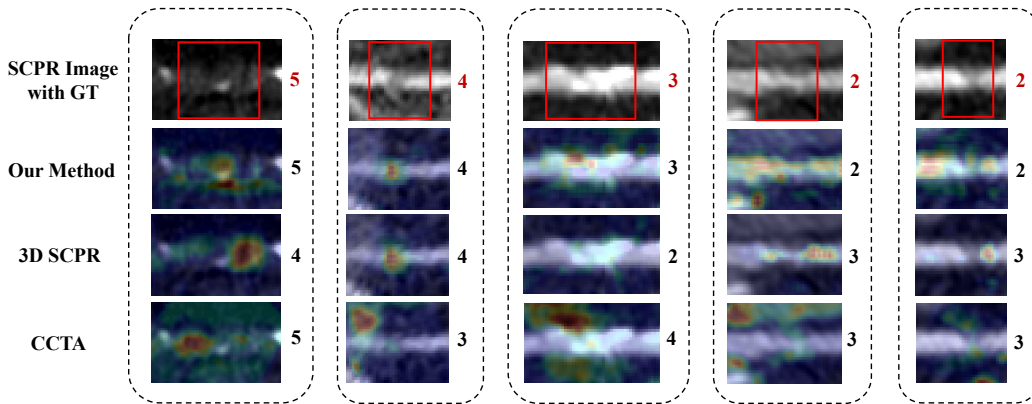

Figure 3: Visual comparison between our approach and other methods. The numerical grade with color of red and black on the right side of each image stands for ground truth and prediction respectively.

Table 2: Ablation study on CFR module.

Data Type		QWK↑	MAE↓	AMAE↓	MCRE↓	AMCRE↓
3D SCPR	CCTA					
✓		0.692	0.470	0.557	1.246	1.300
	✓	0.694	0.506	0.568	1.361	1.215
✓	✓	**0.712**	**0.438**	**0.533**	**1.185**	**1.186**

Table 3: Ablation study on CR Loss.

Metric Function			QWK↑	MAE↓	AMAE↓	MCRE↓	AMCRE↓
AD	SD	CRD					
✓			0.689	0.493	0.569	1.377	1.277
	✓		0.689	0.519	0.564	1.415	1.249
		✓	**0.712**	**0.438**	**0.533**	**1.185**	**1.186**

Qualitative results. To qualitatively explore the effectiveness of the CFR module, we use Grad-CAM(Selvaraju et al., 2017) to generate activation heat maps of the last layer in the backbone network. These heat maps are visualized in Figure 3, and to improve the visualization quality, we project and interpolate the generated 3D activation map into a 2D SCPR image. Since our method utilizes two separate backbones, we average the activation maps from these two backbones to generate a single combined activation map.

Our visualization results indicate that neither 3D SCPR images nor CCTA images can accurately identify the stenosis region on their own. However, the combination of them can effectively learn and highlight the accurate stenosis region. Therefore, our CFR module can improve performance by leveraging the complementary information from both 3D SCPR and CCTA images.

4.3. Ablation Study

Curved Feature Reconstruction. We ablate the effectiveness of different data utilizations in CFR module in Table 2. The results demonstrate that both 3D SCPR images and CCTA images are critical for accurate stenosis grading, and the combination of them can achieve the best performance.

CR Loss. To evaluate the impact of each component in our loss function, we conducted an ablation study on metric function used in Equation (6). The results shown in Table 3 indicate the proposed CR distance can not only perform better in MBPE and AMBPE which are relevant with clinical risk, but also in traditional ordinal regression metrics.

5. Conclusion

In this paper, we propose a novel deep learning-based approach to achieve multi-level stenosis grading for the first time, as far as we know. To address the challenge of predicting stenosis grade accurately in the presence of abnormal distortions in 3D SCPR images, we introduce a novel CFR module that allows for a comprehensive stenosis analysis from both 3D SCPR and CCTA images. Furthermore, we incorporate the clinical relevance of stenosis grading by introducing CR Loss, which encodes the clinical risk boundary into neural network training. Through experiments conducted on a private dataset, we demonstrate that our method outperforms other approaches significantly in both traditional and clinical relevance metrics.

References

Sema Candemir, Richard D White, Mutlu Demirer, Vikash Gupta, Matthew T Bigelow, Luciano M Prevedello, and Barbaros S Erdal. Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary ct angiography with a deep 3-dimensional convolutional neural network. *Computerized Medical Imaging and Graphics*, 83:101721, 2020.

Ricardo C Cury, Jonathon Leipsic, Suhny Abbara, Stephan Achenbach, Daniel Berman, Marcio Bittencourt, Matthew Budoff, Kavitha Chinnaiyan, Andrew D Choi, Brian Ghoshhajra, et al. Cad-rads™ 2.0–2022 coronary artery disease-reporting and data system: An expert consensus document of the society of cardiovascular computed tomography (scct), the american college of cardiology (acc), the american college of radiology (acr), and the north america society of cardiovascular imaging (nasci). *Cardiovascular Imaging*, 15(11):1974–2001, 2022.

Felix Denzinger, Michael Wels, Nishant Ravikumar, Katharina Breininger, Anika Reidelshöfer, Joachim Eckert, Michael Sühling, Axel Schmermund, and Andreas Maier. Coronary artery plaque characterization from ccta scans using deep learning and radiomics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 593–601. Springer, 2019.

Raul Diaz and Amit Marathe. Soft labels for ordinal regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4738–4747, 2019.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Rebekka Vibjerg Jensen, Marie Vognstoft Hjortbak, and Hans Erik Bøtker. Ischemic heart disease: an update. In *Seminars in Nuclear Medicine*, volume 50, pages 195–207. Elsevier, 2020.

B Michael Kelm, Sushil Mittal, Yefeng Zheng, Alexey Tsymbal, Dominik Bernhardt, Fernando Vega-Higuera, S Kevin Zhou, Peter Meer, and Dorin Comaniciu. Detection, grading and classification of coronary stenoses in computed tomography angiography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 25–32. Springer, 2011.

HA Kirişli, Michiel Schaap, CT Metz, AS Dharampal, Willem Bob Meijboom, Stella-Lida Papadopoulou, Admir Dedic, Koen Nieman, Michiel A de Graaf, MFL Meijs, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Medical image analysis*, 17(8):859–876, 2013.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows.

In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Xinghua Ma, Gongning Luo, Wei Wang, and Kuanquan Wang. Transformer network for significant stenosis detection in ccta of coronary arteries. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 516–525. Springer, 2021.

Xinghua Ma, Mingye Zou, Xinyan Fang, Yang Liu, Gongning Luo, Wei Wang, Kuanquan Wang, Zhaowen Qiu, Xin Gao, and Shuo Li. Spatio-temporal contrast network for data-efficient learning of coronary artery disease in coronary CT angiography. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2024 - 27th International Conference, Marrakesh, Morocco, October 6-10, 2024, Proceedings, Part XI*, volume 15011 of *Lecture Notes in Computer Science*, pages 645–655. Springer, 2024. doi: 10.1007/978-3-031-72120-5_60. URL https://doi.org/10.1007/978-3-031-72120-5_60.

Xinghua Ma, Xinyan Fang, Mingye Zou, Gongning Luo, Wei Wang, Kuanquan Wang, Zhaowen Qiu, Xin Gao, and Shuo Li. A trusted lesion-assessment network for interpretable diagnosis of coronary artery disease in coronary CT angiography. In Toby Walsh, Julie Shah, and Zico Kolter, editors, *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 6009–6017. AAAI Press, 2025. doi: 10.1609/AAAI.V39I6.32642. URL https://doi.org/10.1609/aaai.v39i6.32642.

Graham Mowatt, Jonathan A Cook, Graham S Hillis, Shonagh Walker, Cynthia Fraser, Xueli Jia, and Norman Waugh. 64-slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. *Heart*, 94(11):1386–1393, 2008.

Morteza Naghavi, Peter Libby, Erling Falk, S Ward Casscells, Silvio Litovsky, John Rumberger, Juan Jose Badimon, Christodoulos Stefanadis, Pedro Moreno, Gerard Pasterkamp, et al. From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: Part i. *Circulation*, 108(14):1664–1672, 2003.

Kenichiro Otsuka, Shota Fukuda, Atsushi Tanaka, Koki Nakanishi, Haruyuki Taguchi, Junichi Yoshikawa, Kenei Shimada, and Minoru Yoshiyama. Napkin-ring sign on coronary ct angiography for the prediction of acute coronary syndrome. *JACC: Cardiovascular Imaging*, 6(4):448–457, 2013.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

Rahil Shahzad, Hortense Kirişli, Coert Metz, Hui Tang, Michiel Schaap, Lucas van Vliet, Wiro Niessen, and Theo van Walsum. Automatic segmentation, detection and quantification of coronary artery stenoses on cta. *The international journal of cardiovascular imaging*, 29:1847–1859, 2013.

Antonio Tejero-de Pablos, Kaikai Huang, Hiroaki Yamane, Yusuke Kurose, Yusuke Mukuta, Junichi Iho, Youji Tokunaga, Makoto Horie, Keisuke Nishizawa, Yusaku Hayashi, et al. Texture-based classification of significant stenosis in ccta multi-view images of coronary arteries. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 732–740. Springer, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Matthijs J Warrens. Some paradoxical results for the quadratically weighted kappa. *Psychometrika*, 77:315–323, 2012.

Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Yu Zhang, Jun Ma, and Jing Li. Coronary r-cnn: Vessel-wise method for coronary artery lesion detection and analysis in coronary ct angiography. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 207–216, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16437-8.

Majd Zreik, Robbert W Van Hamersvelt, Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. 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.