Direct Quantification of Coronary Artery Stenosis Through Hierarchical Attentive Multi-View Learning

Dong Zhang, Guang Yang, Shu Zhao, Yanping Zhang, Dhanjoo Ghista, Heye Zhang, *Member, IEEE*, and Shuo Li, *Senior Member, IEEE*

***Abstract*— Quantification of coronary artery stenosis on X-ray angiography (XRA) images is of great importance during the intraoperative treatment of coronary artery disease. It serves to quantify the coronary artery stenosis by estimating the clinical morphological indices, which are essential in clinical decision making. However, stenosis quantification is still a challenging task due to the overlapping, diversity and small-size region of the stenosis in the XRA images. While efforts have been devoted to stenosis quantification through low-level features, these methods have difficulty in learning the real mapping from these features to the stenosis indices. These methods are still cumbersome and unreliable for the intraoperative procedures due to their two-phase quantification, which depends on the results of segmentation or reconstruction of the coronary artery. In this work, we are proposing a hierarchical attentive multi-view learning model (HEAL) to achieve a direct quantification of coronary artery stenosis, without the intermediate segmentation or reconstruction. We have designed a multi-view learning model to learn more complementary information of the stenosis from different views. For this purpose, an intra-view hierarchical attentive block is proposed to learn the discriminative information of stenosis. Additionally, a stenosis representation learning module is developed to extract the multi-scale features from the keyframe perspective for considering the clinical workflow. Finally, the morphological indices are directly estimated based on the multi-view feature embedding. Extensive experiment studies on clinical multi-manufacturer dataset consisting of 228 subjects show the superiority of our HEAL against nine comparing methods, including direct quantification methods and multi-view learning methods. The experimental results demonstrate the better clinical agreement between the ground truth and the prediction, which endows our proposed method with a great potential for the efficient intraoperative treatment of coronary artery disease.**

***Index Terms*— direct quantification, X-ray angiography, multi-view learning, intraoperative treatment, coronary artery stenosis.**

Manuscript received June 1, 2020; revised August 9, 2020; accepted August 13, 2020. Date of publication August 17, 2020; date of current version November 30, 2020. This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2019B010110001; in part by the Key Program for International Cooperation Projects of Guangdong Province under Grant 2018A050506031; in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2019B151502031; in part by the National Natural Science Foundation of China under Grant 61771464, Grant U1801265, and Grant U1908211; in part by the Capital Medical Development Research Foundation of China under Grant PXM2020_026272_000013; and in part by the Fundamental Research Funds for the Central Universities under Grant 19lgzd36. *(Corresponding authors: Heye Zhang; Shu Zhao.)*

Dong Zhang and Heye Zhang are with the School of Biomedical Engineering, Sun Yat-sen University, Shenzhen 518001, China (e-mail: zhangd95@mail2.sysu.edu.cn; zhangheye@mail.sysu.edu.cn).

Guang Yang is with the Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, U.K., and also with the National Heart and Lung Institute, Imperial College London, London SW7 2AZ, U.K. (e-mail: g.yang@imperial.ac.uk).

Shu Zhao and Yanping Zhang are with the Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei 230601, China, and also with the School of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: zhaoshuzs2002@hotmail.com; zhangyp2@gmail.com).

Dhanjoo Ghista is with University 2020 Foundation, Northborough, MA 01532 USA (e-mail: d.ghista@gmail.com).

Shuo Li is with the Department of Medical Biophysics, Western University, London, ON N6A 3K7, Canada (e-mail: slishuo@gmail.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.3017275

I. Introduction

QUANTIFICATION of coronary artery stenosis is crucial for the intraoperative treatment of coronary artery disease. The quantification of coronary artery stenosis is carried out to obtain the clinical morphological indices of the stenosis. These indices (namely, lesion length, minimum lumen diameter, and reference vessel diameter, etc., as shown in Fig. 1 (a)) are essential in clinical decision making, regarding the need for coronary artery revascularization [1] and the interventional stent selection. In the clinical setting, X-ray angiography (XRA) is the main imaging methodology to guide the intraoperative treatment by presenting the visualization of the coronary morphology [2]–[5]. However, the clinical stenosis quantification from intraoperative XRA images relies almost entirely on visual estimation or manual measurement [6], by selecting an ideal viewpoint, and then defining a keyframe as necessary. The visual estimation is an inaccurate quantification, which suffers from inter-observer variability [6]. The manual measurement is time-consuming, and it also has variability due to the physician characteristics (e.g., experience, board-certification in cardiology). This subjective quantification inevitably causes unreliable assessment

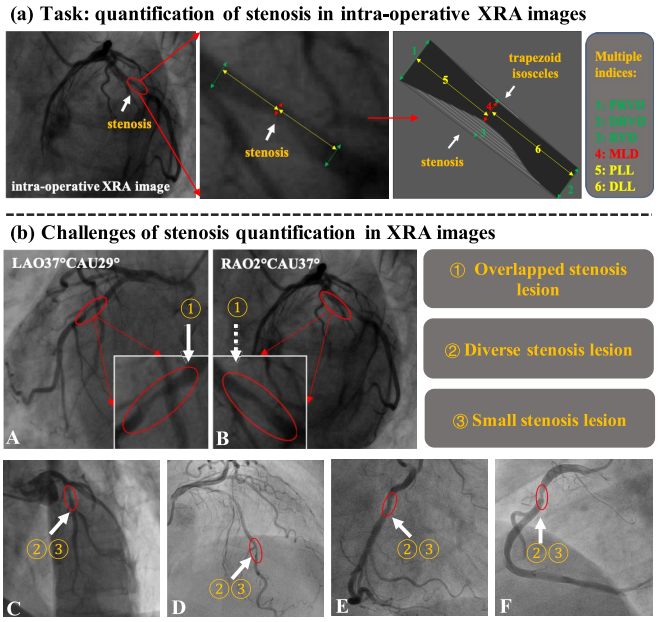

Fig. 1. Quantification of coronary artery stenosis from intraoperative XRA images is a challenging task. (a) The coronary artery stenosis is presented, and the 6 clinical indices including minimum lumen diameter (MLD), proximal and distal reference vessel diameters (PRVD, DRVD), reference vessel diameter (RVD), and proximal and distal lesion length (PLL, DLL) are shown. (b) The overlapping occurs in the stenosis region of the left circumflex coronary artery (LCX) in LAO37°CAU29° viewpoint, compared to the other viewpoint RAO2°CAU37° (A, B). The diverse stenosis lesion is quite small in the entire XRA image, and the representation is difficult to learn (C, D, E, F). Abbreviations: LAO, Left Anterior Oblique; CAU, Caudal; RAO, Right Anterior Oblique.

of the stenosis, which not only decreases the operative workflow efficiency but also raises the health hazards to patients, and even leads to improperly carried out percutaneous coronary intervention (PCI) procedures [7]. Therefore, a reliable computer-aided quantification from XRA images is highly desired in the intraoperative quantification and treatment of stenosis lesion in order to meet the clinical needs, which includes supporting the decision making process, reducing the operative mistakes and improving the clinical workflow efficiency.

However, the coronary artery stenosis quantification from intraoperative XRA images is still challenging due to the following three perspectives shown in Fig. 1 (b). Firstly, the stenoses might overlap with other artery vessel segments in an ideal viewpoint. This overlapping inevitably leads to the information loss of the stenosis morphology. Hence, with only this partial available information, the quantification of the stenosis is less comprehensive and in fact unreliable. Secondly, the stenosis diversities are derived from the large anatomical diversities of the coronary artery stenosis between and within subjects. This makes it difficult to learn the discriminative representation of the stenosis. Thirdly, the stenosis exhibits a large variation in size and is of quite small sizes throughout the entire XRA image. It is hence difficult to map the rare feature information of the small stenosis into the morphology related indices.

Most existing methods [2], [8]–[10] based on the low-level image features make it difficult to tackle the challenges in stenosis quantification from XRA images. They cannot provide the real mapping of the stenosis from the low-level features to the morphological indices, when the overlapping occurs or the morphology of stenosis lesion changes. Even then, these existing methods achieve more reliable quantification accuracy than the visual estimation; but they are still unable to meet the clinical needs of the intraoperative quantification. This is because they are based on the two-phase quantification, by first performing the segmentation or the reconstruction of the coronary artery, and thereafter measuring the clinical indices. Obviously, the two-phase methods suffer from error accumulation and low efficiency in the intraoperative quantification of stenosis. Hence, the stenosis quantification workflow that circumvents intermediate segmentation or reconstruction and performs direct quantification can be instrumental in increasing the efficiency of the intraoperative workflow, and thereby reducing the radiation exposure for patients and cardiologists, as well as decreasing the potential health hazards to patients.

In order to address the aforementioned challenges, we are proposing a **Hi**E**r**archical **A**ttentive mu**L**ti-view learning model, termed **HEAL**, to achieve the direct quantification of coronary artery stenosis. It is designed to mimic the cardiologist in the intraoperative procedure, by performing a comprehensive observation on the stenosis, based on a main viewpoint, a support viewpoint and a keyframe selected from the main viewpoint. In particular, HEAL is comprised of a main-view module, a support-view module, a keyframe-view module, and a regression module, for achieving the direct quantification, as follows: 1. To alleviate the effect brought about by stenosis overlapping, the multi-view learning architecture can provide more complementary information from the two viewpoints of the coronary artery stenosis. 2. Unlike most existing multi-view learning methods that mainly focus on learning consistency and complementarity, our model seeks a discriminative representation for stenosis by extracting the discriminative information existing in consistency and complementarity from two views. The main-view module and the support-view module learn the spatio-temporal features of coronary arteries from the 2D+T image sequence in each view. So then for building the discriminative representation of coronary artery stenosis, we are proposing an intra-view hierarchical attentive block embedded in the main-view module and the support-view module. 3. In order to enhance the representation of the small stenosis, we build a keyframe view for the stenosis from a keyframe perspective, mimicking the clinicians in the intraoperative procedure. Particularly, we introduce several dilated residual blocks with hybrid dilated convolution in order to extract the complementary information with different resolutions for stenosis from the keyframe. Our model also preserves the specific morphological feature from each view, thereby building the view-specific representation. Accordingly, the regression module is aimed to learn the mapping from the comprehensive representation to the quantitative stenosis indices. We have conducted extensive experiments on a clinical XRA dataset to show that HEAL achieves better performance. Specifically, our work contributes to coronary artery stenosis quantification in three different ways:

1. We develop a clinical tool to enable the direct quantification of the coronary artery stenosis in the intraoperative XRA images. This clinical tool can provide the essential guidance for clinicians to analyze and quantify the stenosis from XRA images in the clinical routine.
2. We propose a multi-view learning framework to achieve the stenosis quantification directly. In particular, we propose an intra-view hierarchical attentive block to build the discriminative representation of stenosis, by capturing the pixel region correlation and the intrinsic hierarchical structure of intra view. We develop a stenosis representation learning module to effectively extract the multi-scale semantic features from a keyframe perspective for the clinical workflow. Moreover, this framework can enable the mapping from the feature representation to the indices, as well as improve the adaptability of the index quantification to the stenosis diversity.
3. We experimentally validate that our approach HEAL is able to generate low-error quantification on a multi-manufacturer XRA dataset. The experimental results demonstrate that HEAL compares favorably to existing methods and other direct quantification methods. The effect on the quantification of stenosis existing overlapping is also validated.

This work advances our preliminary work in MICCAI-2019 [11], as follows: (1) A new powerful multi-view learning model now extracts discriminative information for stenosis from XRA images, and it improves the stenosis quantification. (2) An improved keyframe view module has been designed to learn the complementary information with different resolutions, for enhancing the representation of stenosis. (3) A consistency loss has been used to explore the consensus correlations among the multi-view representations. (4) Experiments are extended to a larger multi-manufacturer XRA dataset. (5) More validations with rigorous discussion have been incorporated in the paper.

The remainder of this article is organized as follows. The related works are presented in Section II. The proposed HEAL model is introduced in Section III. Section IV gives detailed descriptions of data acquisition and experimental study. The results are reported and analyzed in Section V. Finally, we provide conclusion in Section VI.

II. Related Works

Multiple studies have been conducted on coronary artery from the XRA images. Most of these works are related to coronary artery segmentation and reconstruction [2], vessel extraction and enhancement [12], [13], and stenosis grading [10]. Only a few works have stressed on the quantification of the coronary artery stenosis. The existing quantification methods cannot directly and accurately quantify the stenosis. They have firstly required complex artery reconstruction [2] or vessel extraction task [10], and then performed a quantitative measurement of the stenosis. Moreover, they cannot provide an effective and reliable guidance in the intraoperative procedures, because of the inefficient and cumbersome workflow. The other quantification methods [14]–[16] for stenosis have used 3D computed tomography angiography (CTA) data; but these methods cannot be directly applied to 2D XRA images in the intraoperative scenarios. Moreover, the direct quantification methods [17]–[19] aim to obtain the relationship between medical images and clinical measurements directly, without the results based on segmentation or reconstruction. These methods have achieved great success recently in many clinical measurement fields, such as multiple cardiac index estimation [18] and spine cobb angle estimation [17]. Furthermore, no attempts have also been made to use them for direct stenosis quantification. The quantification task of coronary stenosis from XRA images has specific challenges for direct methods to be applied to this task.

In this regard, our work on stenosis quantification is also related to the multi-view learning. For many real-world applications, the variables of each data instance can be naturally partitioned into groups. Each variable group is referred to as a particular view, and the multiple views for a particular problem can take different forms, e.g., the medical images from different modalities [20], and different feature types of one object [21]. Due to the effectiveness of exploring the complementarity and consistency among multiple views, the multi-view learning has achieved an impressive performance. Among these works, some unsupervised methods [22], [23] project all views into a latent common space by maximizing the cross-view correlation, typically based on canonical correlation analysis (CCA) [24]. Furthermore, some supervised multi-view learning approaches [25] have been proposed to explicitly exploit the discriminant information such as class labels, by minimizing the between-class correlation and maximizing the within-class correlation. Recently, several deep neural network (DNN) based methods [26], [27] have been proposed to learn high-level common representations from nonlinear correlations across different views. However, the previous works [8], [9] on stenosis quantification have ignored the consistency and the complementarity among multiple viewpoints. Based on the concept of multi-view learning [28], more quantitative vascular information can be obtained by combining the consistent and complementary information among multiple viewpoints.

The XRA images from multiple viewpoints can be used to provide a more descriptive and informative representation for the coronary artery, and to alleviate the effect brought by stenosis overlapping. In the stenosis quantification task, the large anatomical diversities of the coronary arteries between and within subjects are causing immense difficulty in understanding the features of the stenoses via the simple representation of XRA images from several viewpoints. Besides, many multi-view learning methods have been proposed to only explicitly preserve the complementary information of different views [29]. However, not all the complementary information is discriminative, and non-discriminative information still exists in the complementarity [30]. It is hence important to address the discriminative representation for stenosis in multiple views, by exploring the discrimination existing in complementarity, which is the direct factor dominating the learning performance of direct quantification. In clinical angiography procedures, physicians often select a main viewpoint and a support

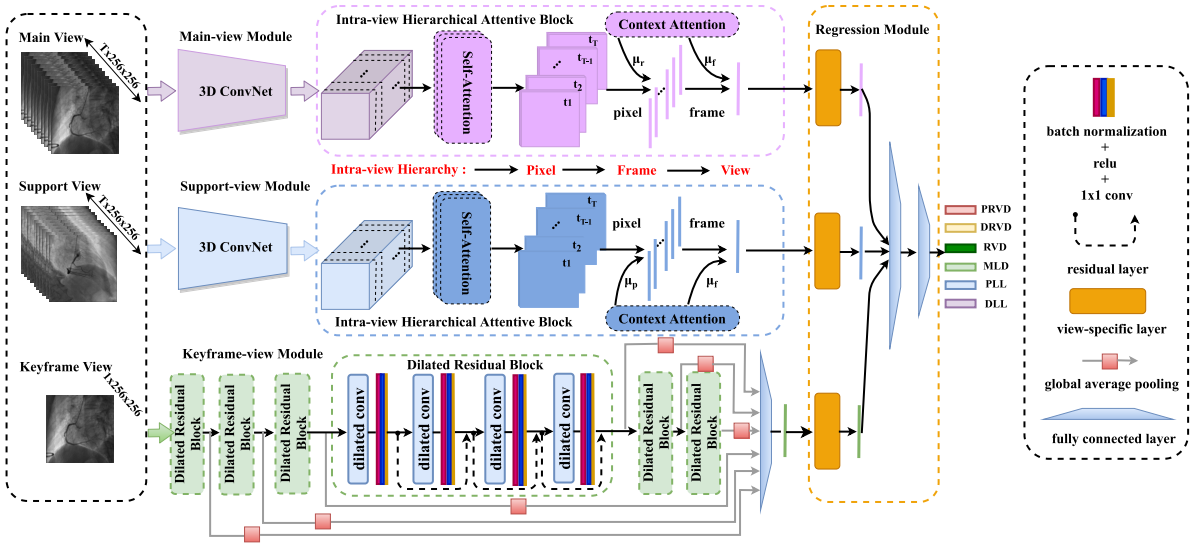

Fig. 2. Framework of HEAL for direct quantification of coronary artery stenosis. It is comprised of three view-modules (main view, support view, keyframe view) and a regression module. In this framework, a 3D ConvNet is utilized to extract comprehensive spatio-temporal feature of coronary artery stenosis from 2D+T images in main view and support view. We propose an intra-view hierarchical attentive block, which extracts the discriminative information existing in consistency and complementarity from main view and support view for the stenosis. Multi-scale features are learned in keyframe-view module to enhance the representation of the stenosis. The regression module fuses the multi-view representation and generates the quantitative indices directly.

viewpoint to observe the stenosis clearly. Hence, we are focusing on the stenosis quantification in a multi-view learning framework.

III. Methodology

Motivated by the clinical needs of the intraoperative scenario, we propose a direct quantification method, HEAL, to estimate the morphological indices for coronary artery stenosis from XRA images. We embed the multi-view learning architecture into a deep network in order to share the complementary information of stenosis from multiple views. HEAL mainly contains three view modules (namely, main view, support view, keyframe view) and one regression module, as shown in Fig. 2. The problem formulation and the details of each module are given below.

A. Problem Formulation

Our direct quantification model HEAL aims to learn the discriminative representation from multiple views, as well as the mapping from the learned representation to the multiple quantitative stenosis indices. It is defined by

$$\begin{aligned} &Given \quad X = \{X_{im}, X_{is}, x_{ikey}\}_{i=1}^{N} \;\; Y = \{Y_i\}_{i=1}^{N} \\ &Objective: \; learn\ the\ mapping \quad f : r(\boldsymbol{X}) \xrightarrow{\Omega} Y \\ &\Omega^* = \arg\min_{\Omega} \; J(heal(r(\boldsymbol{X}); \Omega), Y) \end{aligned} \tag{1}$$

where X, Y are the training examples and labels respectively. $Y_i = \{y_{i1}, y_{i2}, y_{i3}, \ldots, y_{id}\}$ is the label of i_{th} training example. $X_{im}, X_{is} \in (x_1, x_2, \ldots, x_T, \mathbb{R}^{T\times H\times W})$, $x_{ikey} \in \mathbb{R}^{1\times H\times W}$ are the XRA image data from the main view, the support view and the keyframe view of the training sample X_i. In particular, X_{im}, X_{is} are the 2D + T XRA sequential images and x_{ikey} is only the keyframe image in main view. H and W are the height and width of each frame (H = W = 256), T is the temporal step (T = 10). N is the number of training samples, and d is the number of quantitative indices ($d = 6$). The objective of our model HEAL is to learn the mapping $f : r(\boldsymbol{X}) \xrightarrow{\Omega} Y$ from multi-view feature representation $r(\boldsymbol{X})$ of the stenosis to the multiple quantitative indices Y. J is the objective function, where $heal$ indicates the HEAL model, and Ω is the model parameter set to be learned.

B. Multi-View Learning for Comprehensive Representation of Stenosis

1) 3D ConvNet for Spatio-Temporal Feature: Extracting the spatio-temporal feature from the input sequential 2D + T images X_m and X_s in main view and support view is crucial for morphological index estimation. The 3D convolutional neural network (3D ConvNet) has the adequate and good capability to learn the spatio-temporal features, thanks to the operations of 3D convolution. Interestingly, X_m and X_s can also be considered as 3D data ($T \times H \times W$). Therefore, we construct a 3D ConvNet consisting of successive 3D Conv layers to extract the morphology (spatial) and kinematic (temporal) features from different temporal steps for coronary arteries. To preserve the consistent information among the main view and the support view, the 3D ConvNets in the main-view module and the support-view module have the same network architecture, but do not share the same network parameters.

2) Intra-View Hierarchical Attentive Block for Discriminative Representation: The discriminative information existing in consistency and complementarity constitutes actually the direct factor to dominate the learning performance. To extract the discriminative information from each view, we propose an intra-view hierarchical attentive block, which can capture the

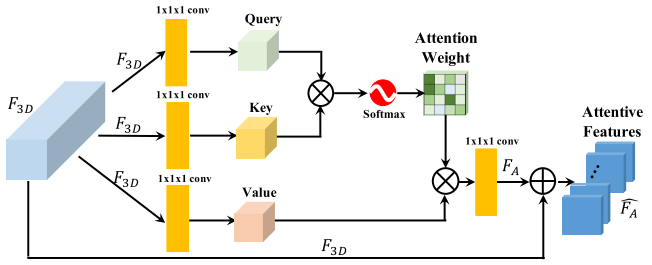

Fig. 3. Illustration of the self-attention module in our proposed intra-view hierarchical attentive block. The input and output are a $T \times C \times M$ feature map. $\otimes$ denotes the batch matrix multiplication and $\oplus$ denotes element-wise summation.

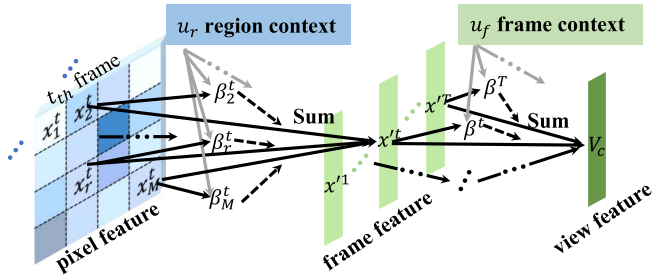

Fig. 4. Illustration of the context attention module in our proposed intra-view hierarchical attentive block.

pixel correlation and the intrinsic hierarchical structure of sequence-shaped data. This structure information is a two-level hierarchical information, consisting of the information from the pixel region to the frame, as well as the information from the frame to the view. This intra-view hierarchical attentive block is comprised of a self-attention module, and a context-attention module consisting of a pixel-context attention layer and a frame-context attention layer.

Inspired by the success of the self-attention mechanism [31], [32], our self-attention module learns the discriminative representation of the stenosis by capturing the interaction correlations among the image pixel regions, as shown in Fig. 3. In particular, given the intra-view frame sequence feature $F_{3D} \in \mathbb{R}^{T\times C\times M}$ learned from 3D ConvNet, the feature map F_{3D} is first transformed into queries Q, keys K, and values V matrices by different linear projections. Here, C is the number of channels and M is the number of the pixel regions of the t_{th} frame feature map $x^t \in \mathbb{R}^{C\times M}$ of F_{3D}. For the t_{th} frame feature map x^t, the corresponding transformations are $Q^t = W_q^t x^t$, $K^t = W_k^t x^t$, $V^t = W_v^t x^t$, where $W_q^t \in \mathbb{R}^{C/8\times C}$, $W_k^t \in \mathbb{R}^{C/8\times C}$, and $W_v^t \in \mathbb{R}^{C/8\times C}$ are the learned weight matrices.

$$\begin{aligned} s_{ij}^t &= {Q_i^t}^\top K_j^t \\ \alpha_{j,i}^t &= \frac{exp(s_{ij}^t)}{\sum_{i=1}^{M} exp(s_{ij}^t)} \end{aligned} \tag{2}$$

In the above formulation, $\alpha_{j,i}^t$ is the self-attention weight, which indicates the interaction correlation between the i_{th} pixel region and the j_{th} pixel region. Next, the attentive feature of the stenosis is computed, based on the following equation:

$$A_j^t = W_g^t(\sum_{i=1}^{M} \alpha_{j,i} V_i^t) \tag{3}$$

where the $W_g^t \in \mathbb{R}^{C\times C/8}$ is the learned weight matrix. All the attentive feature matrices produced by the Eq. (3) are concatenated together and then reshaped to form a unifying matrix A^t. For the attentive feature, the T-frame sequence is denoted as F_A. Finally, to get the comprehensive correlation feature map, a residual-like connection layer with a weighted attention feature map is added into the self-attention module. The detailed computations are based on the following Eq. (4), where γ is a scale factor which can be learned automatically.

$$\begin{aligned} A^t &= concat(A_1^t, A_2^t, \ldots, A_M^t) \\ F_A &= concat(A^1, A^2, \ldots, A^t, \ldots, A^T) \\ \widehat{F_A} &= \gamma\, F_A + F_{3D} \end{aligned} \tag{4}$$

However, not all the regions contribute equally to the representation of each frame meaning. Hence, we introduce the pixel-context attention, as shown in Fig. 4, to extract such pixel regions that are important to the meaning of the current frame. Then we aggregate the representation of these informative pixel regions to form an expressive feature vector of the frame. That is, we first feed the r_{th} region x_r^t of the t_{th} frame through a one-layer MLP to get u_r^t as the hidden feature of x_r^t. Then we measure the importance of the pixel region with a region-level context vector u_r, and we get a normalized importance weight β_r^t through a softmax function. After that, we compute the t_{th} frame representation x'^t as a weighted sum of the regions based on β_r^t.

$$\begin{aligned} u_r^t &= tanh(W_r x_r^t + b_r^t) \\ \beta_r^t &= \frac{exp({u_r^t}^\top u_r)}{\sum_r^M exp({u_r^t}^\top u_r)} \\ x'^t &= \sum_r^M \beta_r^t x_r^t \end{aligned} \tag{5}$$

To reward the frames that are important to the current view representation, we again use the context attention mechanism and introduce a frame-level context vector u_f to measure the importance β^t of the t_{th} frame. Finally, the vector V_c denotes the view representation (i.e. the main view or the support view) that summarizes the information of T frames in the current view. This yields:

$$\begin{aligned} u^t &= tanh(W^t x'^t + b^t) \\ \beta^t &= \frac{exp({u^t}^\top u_f)}{\sum_t^T exp({u^t}^\top u_f)} \\ V_c &= \sum_t^T \beta^t x'^t \end{aligned} \tag{6}$$

3) Keyframe-View Module for Representation Enhancement: The keyframe-view module is designed to enhance the representation of the stenosis from a particular keyframe perspective. This module is motivated by the intraoperative scenario,

in which the cardiologists also have a targeted observation of the stenosis in a keyframe image x_{key}. The commonly-used convolution neural networks are unsuitable to extract the features for the small stenosis lesions in the XRA images, because it is difficult to preserve the detailed information of the small object during the down-sampling process. Thus, the dilated residual block (DRB) is introduced to preserve the more detailed information of the stenosis, which contains useful information for the stenosis index estimation. However, in the standard dilated convolution framework, there exists the "gridding issue" [33]. So we alleviate the gridding effect by introducing the hybrid dilated convolution (HDC) [34] into our dilated residual block. The dilated convolution in DRB can increase the receptive field by enlarging the spatial gap between the sampling points for convolution. However, the high-level semantic feature maps produced by DRBs lack the detailed information of the stenosis in the low-level semantic features, which contains the useful information related to stenosis indices. In order to preserve the detailed information of stenosis, the keyframe-view module fuses the feature information from different levels into a multi-scale feature vector. This characteristic enables our keyframe-view module to simultaneously preserve the semantic features and the local details of the stenosis in the keyframe perspective. In particular, the global average pooling is applied to the different level features generated by each DRB. It sums up the spatial information of the input feature maps, which are transformed into a fixed low-dimensional feature vector. To this end, the fused multi-scale feature V_{key} of the stenosis is developed from the keyframe view.

C. Regression Module for Multiple Stenosis Index Estimation

The regression module aims to aggregate the multi-view feature representation of stenosis for the direct index estimation. However, the concatenated feature representation from different views can cause the over-fitting on a small training sample [28], because the specific statistical property of each view is ignored. Therefore, a view-specific layer proposed in [27] is embedded in the regression module to consider the specific property of each view. We denote the $V_v \in \{V_m, V_s, V_{key}\}$ as the extracted features V_m, V_s, V_{key} from the main-view module, the support-view module and the keyframe-view module. The output of the specific layer is denoted as $S_v \in \{S_m, S_s, S_{key}\}$, which is defined as

$$S_v = V_v \odot sigmoid(log(abs(V_v))) \tag{7}$$

where the function $sigmoid(log(abs(V_v)))$ and S_v are the specific property scores and the view-specific feature of the corresponding view, respectively. In the Eq. (7), $\odot$ is the Hadamard product, and the $abs(\cdot)$ is the absolute value function.

In order to map the learned multi-view features to the stenosis indices, the view-specific features S_m, S_s, S_{key} are concatenated, and then they are fed into a two-layer fully connected network. The output of the regression module is $f(x_i) = W_o(S_m \oplus S_s \oplus S_{key}) + b_o$, where the $\oplus$ denotes the concatenation operator, and W_o and b_o are the weight matrix and bias, respectively. The objective function is shown as Eq. (8):

$$J = \frac{1}{N}\sum_{i=1}^{N}(|f(x_i) - Y_i| + \eta_{con}\sum_{a\neq b}(V_{ia} - V_{ib})^2 + \lambda_{qca}||w_i||^2) \tag{8}$$

In this formulation, to minimize the difference between the outputs $f(x_i)$ and the ground truth, we employ the mean absolute error (MAE) as the loss function. The second term in Eq. (8) is used to explore the consensus correlations among multiple views, where $a, b \in \{m, s, key\}$, and η_{con} is the weight parameter to consider the consensus correlations.

D. Model Configuration

Our proposed direct quantification method HEAL consists of a main-view module, a support-view module, a keyframe-view module and a regression module. The main-view module and the support-view module both have a 3D ConvNet for spatio-temporal features of the coronary artery stenosis, and an intra-view hierarchical attentive block for discriminative representation of stenosis. In particular, the 3D ConvNet consists of 6 3D convolutional layers. Each convolutional layer is followed by a batch normalization layer and a leaky rectified linear unit. We use 64, 128, 128, 256, 256, 512 convolutional kernels for these 6 layers, respectively. The first 3 layers have $2 \times 3 \times 3$-sized kernels, and the last 3 layers have $2 \times 2 \times 2$-sized kernels. We use the $1 \times 2 \times 2$ stride in each convolutional layer without the max-pooling operation. In the intra-view hierarchical attentive block, the learned weight matrices W_q^t, W_k^t, W_v^t and W_g^t in self-attention module are all implemented as $1\times1\times1$ convolutions, and the scalar parameter γ is initialized as 0 in training. In the context attention module, the region-level context vector u_r and the frame-level context vector u_f are both initialized as a C-dimensional vector, where C is the number of channels of the frame feature maps.

The keyframe-view module is comprised of 6 dilated residual blocks. Each dilated residual block has one 3×3 stride $(s = 2)$ convolution layer and a hybrid dilated convolution group, which consists of three succeeding 3×3 dilated convolutional layers with their dilation rates as 1, 2 and 3, respectively. The numbers of convolutional kernels of the 6 dilated residual blocks are 32, 64, 128, 256, 512, 512, respectively. Each convolutional layer is followed by a batch normalization layer and a leaky rectified linear unit. To capture the multi-scale features of the stenosis, each dilated residual block is connected with a global average pooling layer. A 1504-dimensional feature vector is composed by concatenating the pooled features from different feature levels. A fully connected layer with 512 units is then used to form a keyframe-view representation based on the multi-scale features. In the regression module, two fully connected layers with 512, 6 units are applied to estimate the morphological indices.

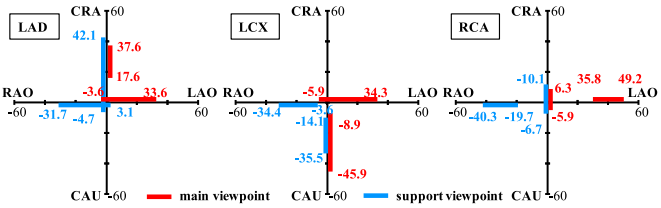

Fig. 5. Acquisition angles of two viewpoints for LAD, LCX, RCA. Abbreviations: LAO, Left Anterior Oblique; RAO, Right Anterior Oblique; CRA, Cranial; CAU, Caudal.

IV. Experimental Study

A. Data Acquisition

This study includes angiographic images retrospectively collected clinically from 228 subjects in First People's Hospital of Shunde and Sun Yat-Sen Cardiovascular Hospital of Shenzhen. The angiographic images were obtained at 15 frames/s by monoplane x-ray systems from 3 manufacturers (200 subjects from Philips Allura Xper, 15 subjects from GE Healthcare Innova, 13 subjects from Siemens AXIOM-Artis). The subjects' ages were from 31 yrs to 87 yrs, with an average of 63.6 yrs. The size of the XRA images in our dataset is 512×512, and the pixel spacing ranges from 0.183 mm/pixel to 0.741 mm/pixel, with the mode of 0.37 mm/pixel. All subjects had the approval of the institutional review board for a retrospective observational study. The X-ray coronary angiography series comprised more than 2 acquisitions per major vessel (right coronary artery (RCA), left anterior descending artery (LAD), left circumflex artery (LCX) at different viewpoints. An LAO (Left Anterior Oblique)-based acquisition viewpoint and an RAO (Right Anterior Oblique)-based acquisition viewpoint were selected as the main viewpoint and the support viewpoint, in which the stenosis region could be observed clearly. The viewpoint angles for different coronary arteries were acquired from certain angulation ranges, which are presented in Fig. 5. The XRA image sequences of the two acquisition viewpoints start from the end-diastolic angiographic frame to the subsequent 10_{th} frame. The XRA image sequences cover a complete cardiac cycle. The keyframe image is defined as the frame in the end-diastole phase of the main viewpoint [35]. In particular, in order to avoid the uncertainty of the quantification, the keyframe is selected by the physicians in our work. The ground truth values were manually measured independently by two experienced (over 10 years) medical physicians [36], [37] (The interobserver variabilities are 0.09 mm, 0.08 mm, 0.1 mm, 0.07 mm, 0.22 mm, 0.25 mm for PRVD, DRVD, RVD, MLD, PLL, DLL, respectively).

B. Experimental Settings and Evaluation Metrics

For training the HEAL model, we optimize it by using the Adam optimizer with a batch of 16 subjects per step and an initial learning rate of 0.0002. The decay rate is set to 0.95, the hyper-parameters η_{con} and λ_{qca} are set to 10^{-2}, 10^{-6}, respectively. In our experiments, 10-fold cross-validation (no images from the same patient are in different groups) is employed on the 200-subject dataset from Philips Medical System. Moreover, the XRA image data from another 2 manufacturers (GE Healthcare and Siemens) are used to evaluate the robustness of HEAL to the difference in manufacturers. Pixel values are normalized to [0, 1] in all the XRA images, which are also resized into 256×256. All the values of the ground truth have been normalized. Our model was implemented by using TensorFlow 1.11.0, and it was trained and tested on an NVIDIA Tesla P40 24GB GPU.

The metrics for evaluating the performance of approaches include Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC). The MAE measures the consistency of prediction results and real values. The PCC measures the linear relationship between the predicted vector $\hat{y}_i$ and the target vector y_i. We also adopt a clinical index stenosis grading accuracy (ACC) to evaluate the performance of our method.

C. Experiments

Extensive experiments were conducted to validate the effectiveness of our proposed method HEAL from the following aspects.

1) Coronary Artery Stenosis Quantification: The performance of HEAL for the quantification of the coronary artery stenosis is examined on our dataset. To evaluate the quantification accuracy and the effectiveness of HEAL, we further compare HEAL with two baseline methods CNNs [38] and 3DCNNs [39], and four direct quantification methods, including HOG+RF [40], Indices-Net [18], DMTRL [41], DMQCA [11], and two multi-view learning methods MVCNN [26], GVCNN [27], as well as an existing coronary vessel diameter estimation method (denoted as Wan *et al.*) [10]. Besides, we have graded the subjects into three categories of Mild, Moderate and Severe (107, 80, and 13, respectively), according to the quantitative coronary arterial grading as recommended by the Society of Cardiovascular Computed Tomography [42]. Additionally, the stenosis grading ability of our framework is evaluated with the comparing methods. It is to be noted that CNNs, HOG+RF, Wan *et al.* and Indices-Net are performed on all the keyframe data. Then, 3DCNNs, DMTRL, DMQCA, MVCNN and GVCNN are performed on the multi-view XRA data.

2) Benefit of Multi-View Learning Framework: The benefit of multi-view learning framework is made evident. Multi-view learning for stenosis quantification is capable of exploiting the complementarity among two viewpoints, and this framework makes it alleviate the effect of stenosis overlapping. To demonstrate this, the quantification performances of the multi-view framework and different combinations of different view modules are compared in terms of stenosis index estimation: i) HEAL, which is our proposed multi-view framework; ii) Single-view models, including Main-view model, Sup-view model and Key-view model, which indicate the quantification frameworks from the main viewpoint, the support viewpoint and the keyframe view, respectively; iii) Two-view models, including Main+Sup model, Main+Key model and Sup+Key model, which indicate the three combinations of each two views. In the Main-view model and Sup-view model, one fully-connected layer is employed, followed by the intra-view hierarchical attentive block to directly estimate the stenosis indices based on the single-view feature representation.

TABLE I
PERFORMANCE COMPARISONS OF THE PROPOSED METHOD HEAL WITH TWO BASELINE METHODS, FOUR EXISTING DIRECT METHODS, TWO MULTI-VIEW LEARNING METHODS AND A CORONARY ARTERY DIAMETER ESTIMATION METHOD. MAE (mm), PCC (%) AND ACC (%) ARE ILLUSTRATED. THE SUPERSCRIPT SYMBOL IN METHOD COLUMN INDICATES THE NUMBER OF VIEWS

Method	View	PRVD	DRVD	RVD	MLD	PLL	DLL	MAE ↓	PCC ↑	ACC ↑
CNNs [38]	Key	0.8964	0.9421	0.9454	0.7387	2.7085	2.7711	1.5004 ± 0.7264	85.36 ± 15.63	69.50 ± 10.55
HOG+RF [40]	Key	0.7077	0.7232	0.6612	0.5683	2.7660	2.8042	1.3718 ± 0.6689	90.45 ± 11.65	73.00 ± 9.876
Indice-Net [18]	Key	0.9188	0.9765	0.8830	0.8869	2.8181	2.7506	1.5390 ± 0.7024	86.89 ± 15.49	72.50 ± 9.042
Wan et al.[1] [10]	Key	0.6925	0.7071	0.6498	0.5643	2.8311	2.6184	1.3438 ± 0.6517	90.81 ± 11.88	79.00 ± 6.849
3DCNNs[1] [39]	Main	0.6968	0.7406	0.6689	0.6453	2.6025	2.6339	1.3313 ± 0.6608	90.48 ± 11.69	73.50 ± 11.95
3DCNNs[1] [39]	Sup	0.7318	0.7845	0.7118	0.6167	**2.4864**	2.5508	1.3137 ± 0.6306	90.75 ± 11.23	71.00 ± 12.11
3DCNNs[2] [39]	Main, Sup	0.7179	0.7559	0.6955	0.6478	2.5689	2.5217	1.3180 ± 0.6570	90.80 ± 11.63	74.00 ± 9.501
DMTRL[1] [41]	Main	0.7267	0.8054	0.7124	0.5683	2.4946	2.5245	1.3053 ± 0.6627	90.45 ± 11.24	71.00 ± 7.775
DMTRL[1] [41]	Sup	0.6844	0.7897	0.7121	0.6340	2.5013	2.5436	1.3064 ± 0.6265	90.03 ± 10.32	72.50 ± 7.992
DMTRL[2] [41]	Main, Sup	0.7632	0.7763	0.7121	0.6372	2.5115	2.5385	1.3231 ± 0.6301	90.55 ± 10.70	72.00 ± 6.939
MVCNN[2] [26]	Main, Sup	0.8436	0.8365	0.8232	0.6764	2.7243	2.4390	1.3905 ± 0.6720	90.10 ± 11.52	70.00 ± 7.026
GVCNN[3] [27]	Main, Sup, Key	0.7384	0.7293	0.7085	0.5573	2.6835	2.6405	1.3428 ± 0.6608	91.12 ± 11.23	74.50 ± 8.357
DMQCA[3] [11]	Main, Sup, Key	**0.6706**	0.7091	0.6237	0.5849	2.5194	**2.4916**	1.2666 ± 0.6357	90.70 ± 11.96	81.50 ± 7.192
HEAL	Main, Sup, Key	0.6800	**0.6506**	**0.6156**	**0.5514**	2.5087	2.4926	**1.2498 ± 0.6487**	**91.32 ± 11.01**	**85.00 ± 8.109**

We also discuss the quantification performance of the stenoses existing overlapping in our multi-view learning architecture. We present the quantification error of the subjects associated with stenosis overlapping in the test data of each fold experiment.

3) Benefit of Intra-View Hierarchical Attentive Block: To investigate the contribution of the intra-view hierarchical attentive block in our proposed model, we conduct comparison experiments between the models with this hierarchical block and the models that remove it. Clinically, the intraoperative decision making always depends on the diagnosis at the artery level (LAD, LCX and RCA, 66, 48 and 86, respectively) and stenosis grade level (Mild, Moderate, Severe). Therefore, we additionally evaluate the quantification ability of our proposed method at both the artery level and the stenosis grade level.

4) Effectiveness of the Keyframe-View Module: We evaluate the effectiveness of the developed keyframe-view module. The keyframe-view module can enhance the expressiveness of the stenosis representation, and thus make it more compatible with the regression module. This is evidenced by comparing the performance of HEAL and that without the keyframe module. We also extract the three representations: i) F_{HEAL}, which is obtained from the HEAL model; ii) F_{nokey}, which is obtained from the HEAL model without the keyframe-view module; iii) $F_{\underline{HEAL}}$, which is obtained from the HEAL model without the intra-view hierarchical attentive block. Each of them is a feature vector of length 512 from the last full-connected layer of the corresponding model. Once the three representations are available, the Random Forest Regressor models with the same configuration are applied to them for stenosis quantification following the ten-fold cross validation protocol. As part of the analyses, we also investigate the mutual information [43] of each feature representation and the ground truth.

5) Robustness, Computation and Parameter Influence: The robustness of our model is evaluated on the XRA images from the other two manufacturers (i.e. GE Healthcare and Siemens), compared with the nine existing methods. The clinical facilitation and effectiveness are also demonstrated by comparing the processing time of our model with the frame rate of XRA sequences in the PCI procedure. Additionally, the effects of the two hyper-parameters (η_{con} and λ_{qca}) are also discussed.

V. RESULTS AND ANALYSIS

This section presents the results and analyses of our experimental study. Section V-A presents the performance of our HEAL on the clinical dataset. Section V-B shows the effectiveness of the multi-view learning framework, compared with the different combinations of the different view modules. Section V-C shows the benefit of the intra-view hierarchical attentive block in HEAL. Section V-D tries to identify whether the keyframe view module can enhance the expressiveness of stenosis. Section V-E mainly analyzes the robustness of the HEAL model to different XRA acquisition manufacturers. The computation analysis and the influences of parameters are presented in Section V-F and Section V-G. Finally, the limitation of HEAL is presented in Section V-H.

A. Coronary Artery Stenosis Quantification

Our proposed HEAL method achieves accurate estimation of the stenosis indices, as shown in the last row of Table I. The MAE values of the six indices (PRVD, DRVD, RVD, MLD, PLL, DLL) are 0.6800, 0.6506, 0.6156, 0.5514, 2.4926, 2.4926, respectively. HEAL estimates the stenosis indices with an average MAE of 1.2498 ± 0.6487 mm, which is around 3 times the mode pixel spacing (0.37 mm/pixel) of the XRA images (512×512) in our dataset. HEAL also achieves the best average PCC correlation of 91.32% with the ground truth. In particular, our model can achieve an 85.00% classification accuracy of stenosis grading using the estimated values.

TABLE II
CONFUSION MATRIX OF CLASSIFICATION RESULTS OF OUR METHOD IN TERMS OF STENOSIS GRADING

Stenosis grading	Mild	Moderate	Severe
Mild	95	12	0
Moderate	13	67	0
Severe	3	2	8

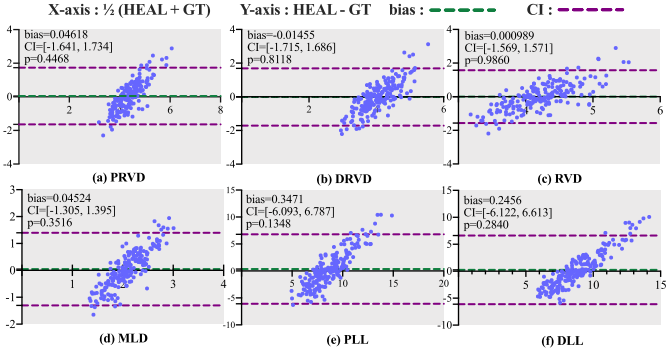

Fig. 6. High agreement between our HEAL and the ground truth (GT) assessed by the Bland-Altman analysis method with respect to 6 clinical indices. The purple dashed lines indicate the 95% confidence intervals of the bias. The green dashed lines indicate the bias. p indicates the statistical hypothesis testing (p-value) between the ground truth and estimated indices.

A confusion matrix is presented in Table II. The results indicate that our model performs better on Mild and Moderate subjects than the Severe subjects. Especially about 40% of severe stenoses are mistakenly graded as Mild and Moderate. However, in this work, only 13 severe stenoses were investigated. Future work may integrate diagnostic knowledge from other modalities imaging data. Nevertheless, a larger training set of subjects with severe stenoses would be required.

Fig. 6 illustrates that the estimated results of HEAL highly agree with the ground truth. Each subplot shows the results of the Bland-Altman analysis for a clinical index. The Y-axis indicates the bias between the values of this clinical index estimated by HEAL and the ground truth. The X-axis indicates the average of these two values. These results indicate the clinical agreement between the HEAL estimation and the ground truth. Moreover, the statistical hypothesis testing (p-value, denoted as p in Fig. 6) between the ground truth and estimated indices are calculated to evaluate the quantification accuracy. No significant differences ($p > 0.05$) were found between the quantification results obtained using our HEAL and the physician-measured values in the XRA images. Our HEAL can also work for different stenosis lesion types. The average quantification errors of our model are 1.4915 ± 0.4714 mm, 0.9506 ± 0.4259 mm, and 1.9871 ± 0.7031 mm for 36 Type A, 126 Type B, and 38 Type C lesions in our dataset, respectively.

The effectiveness of HEAL is shown in Table I by comparisons with CNNs [38], 3DCNNs [39], four direct quantification methods [11], [18], [40], [41], two deep neural network based multi-view learning methods MVCNN [26], GVCNN [27], as well as a coronary vessel diameter estimation method Wan *et al.* [10]. It can be seen that HEAL outperforms the existing direct methods, including our preliminary method [11], with clear MAE reductions and correlation improvements for the clinical indices of stenosis. This makes it evident that our proposed method HEAL can appropriately learn the mapping from the extracted features to the stenosis indices. This is due to the expressive feature representation, which integrates the discriminative complementary information from multiple views of the stenosis. This work exceeds our previous work DMQCA [11] through the more discriminative representation extracted by a more effective multi-view learning framework. The average MAE reductions of HEAL over Wan *et al.* [10] are 6.92% for the stenosis indices. This provides evidence that the HEAL stenosis features are more expressive than the low-level image features in Wan *et al.* method.

B. Benefit of Multi-View Learning Framework

As demonstrated by the experiment results in Fig. 7, our multi-view learning framework can significantly improve the stenosis quantification accuracy, when being compared with the different combinations of different views (Main view, Support view, Keyframe view). As shown in Fig. 7, the models removing any of the single-view modules from the multi-view framework (HEAL) reduce the performance of stenosis quantification. The multi-view model HEAL can obtain the best quantification accuracy. This is because HEAL learns to mimic the reporting clinicians in the intraoperative scenario, to perform a comprehensive observation on the stenosis based on the morphological information from multiple views. However, the sub-frameworks (Main, Sup and Key) consisting of only one view module fail to achieve accurate quantification. This is because the single-view models lack the more comprehensive feature representation from the XRA image data in only one view. Moreover, the sub-frameworks consisting of two view modules have the better quantification performance compared to the single-view models, by integrating the more complementary information of the stenosis from another view.

To further analyze the benefit of fusing the different complementary information from multiple views, we now show the performance of our proposed method HEAL and different sub-frameworks for the alleviation of the influence of the stenosis overlapping on the stenosis quantification. From Fig. 8, it can be seen that our model HEAL outperforms the models using only one view or two views. Besides, these results also show that all the three single-view models (i.e., Main, Sup, Key) achieve lesser quantification performance on the subjects associated with stenosis overlapping than the multi-view models. In particular, the multi-view frameworks (i.e., HEAL, Main+Sup, Main+Key, and Sup+Key) achieve lower MAE values as 1.2190, 1.1989, 1.2450, 1.2537, respectively, than the single-view frameworks. HEAL obtains the best PCC (91.09%) on these subjects. The frameworks Main+Sup, Main+Key also achieve better PCC values (90.53%, 90.65%) than the single-view frameworks. The reason could be that the features related to the morphological information are relatively less from only one view, which is

	PRVD	DRVD	RVD	MLD	PLL	DLL	Ave MAE (mm)	PCC (%)	ACC(%)
HEAL[1] (Main)	0.7006	0.7340	0.7185	0.6423	2.5208	2.5795	1.3160	90.68	76.00
HEAL[1] (Sup)	0.7051	0.7051	0.6366	0.6195	2.6664	2.4878	1.3043	90.10	81.00
HEAL[1] (Key)	0.6934	0.6859	0.6270	0.5605	2.6031	2.6036	1.2956	91.19	83.00
HEAL[2] (Main, Sup)	0.7215	0.6825	0.6700	0.6081	2.5305	2.5010	1.2856	90.51	83.00
HEAL[2] (Main, Key)	0.7016	0.6805	0.6321	0.5692	2.5386	2.5200	1.2737	91.17	83.00
HEAL[2] (Sup, Key)	0.6817	0.6909	0.6376	0.5717	2.6133	2.5913	1.2978	90.99	80.50
HEAL (Main, Sup, Key)	0.6800	0.6506	0.6156	0.5514	2.5089	2.4926	1.2498	91.32	85.00

Higher error

Lower error

Fig. 7. Performance comparison of the proposed multi-view framework HEAL with different combinations of different view modules (Main view, Support view, Keyframe view). 3 single-view models and 3 two-view models are presented. The superscript symbol on the left indicates the number of views.

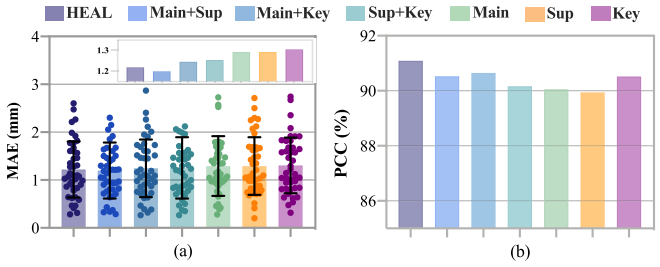

Fig. 8. MAE (a) and PCC (b) of our proposed multi-view model and the single-view/two-view models on the subjects associated with stenosis overlapping. Three single-view models and three two-view models are used in this group of comparisons. The error bars in (a) denote the standard deviations of results. The sub-figure in (a) is the local enlarging display for the MAE values.

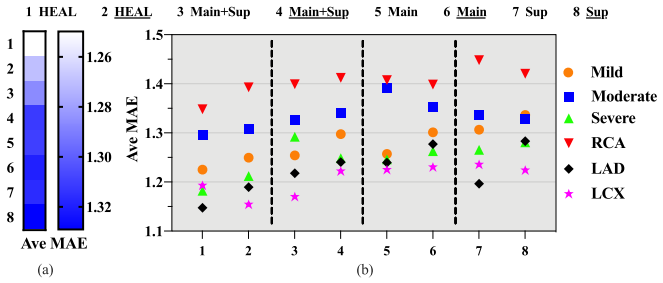

Fig. 9. The benefits of the proposed IHAB. Numbers **1-8** indicate the different models, of which the underlined models indicate the models that remove the IHAB. (a) Average MAE of 8 different models. (b) Quantification on subjects with different types of stenosis grade and coronary arteries.

not enough to achieve a better quantification for the stenosis. However, the multi-view model can integrate more complementary information from other views, which benefits to learn the mapping from the comprehensive features to the stenosis indices.

C. Benefit of Intra-View Hierarchical Attentive Block

As observed in Fig. 9, the models with intra-view hierarchical attentive block (IHAB) can improve the quantification accuracy compared with the models that remove IHAB. For convenience of description, we name this block as IHAB. The heatmap Fig. 9 (a) shows that the models (**1**, **3**, **5**, **7**) with IHAB have lower quantification errors compared to the corresponding models (**2**, **4**, **6**, **8**) removing the IHAB. The lighter-colored squares depict the lower the MAE values. This indicates that the IHAB can learn the more expressive representations correlated with the clinical indices, by integrating the discriminative information of stenosis in a view.

To further evaluate the clinical benefits of the proposed IHAB for stenosis quantification, we analyze the results achieved on the artery level and the stenosis grade level, as shown in Fig. 9 (b). In brief, HEAL achieves better quantification performance on different types of coronary arteries (1.1475 mm for LAD, 1.1932 mm for LCX, 1.3481 mm for RCA), and different degrees of stenosis (1.2250 mm for Mild, 1.2957 mm for Moderate, 1.1821 mm for Severe). Overall, the obtained results (**1** vs **2**, **3** vs **4**, **5** vs **6**, **7** vs **8**) show that the models integrating the IHAB can achieve lower quantification error than the models removing the IHAB.

On the artery level, the models with IHAB have superior effects on RCA and LAD, as indicated by the red inverted triangle and black diamond, respectively. At the stenosis grade level, the models with IHAB perform well on the mild stenosis and severe stenosis, as indicated by the orange circle and green upper triangle. However, the pixel regions around the stenosis are different in XRA images. This demonstrates that the models integrating IHAB do have the ability to learn the discriminative information of stenosis, by capturing the interaction relationships among the different pixel regions. Moreover, we visualize the discriminative representation that plays a vital role in stenosis quantification. Fig. 10 (a)-(1)∼(4) present representative examples of different stenosis lesion types (A, B, C) and the normal angiogram to show the discriminative features learned by the IHAB. From left to right, each column indicates the original XRA images, the features extracted by HEAL with IHAB and the model without IHAB, respectively. The features learned by IHAB focus on coronary artery stenosis regions, which contributes to the accurate quantification. Moreover, Fig. 10 (a)-(4) indicates that the IHAB pays more attention to the coronary artery region in the XRA image. Additionally, an angiogram instance with multiple stenoses is also presented in Fig. 10 (b)-(5), which indicates our model can extract more discriminative features for the severer one (stenosis B). This is consistent with the real PCI scenario, in which a stent placement is performed for the severer one first.

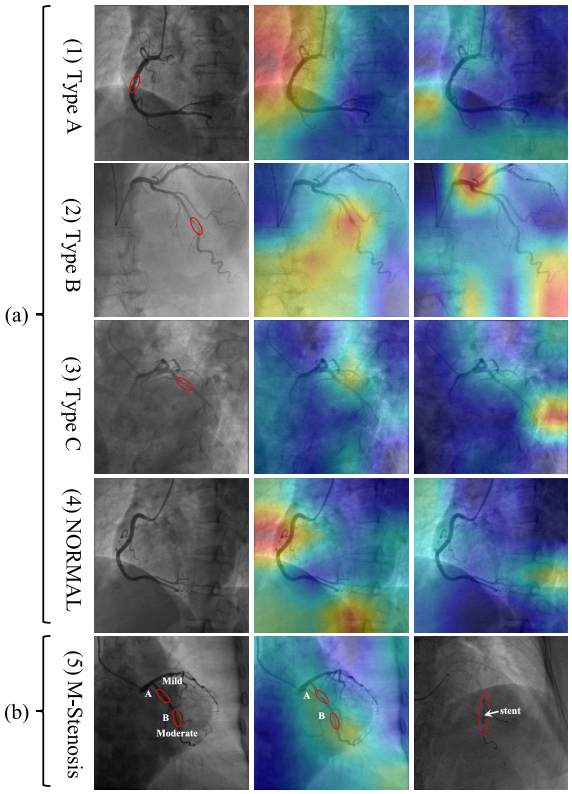

Fig. 10. Discriminative feature representation learned from IHAB module. (a) Representative subjects of different lesion types (Type A, B, C) and the normal (NORMAL) angiogram are visualized. From left to right, each column indicates the original images, the features learned by models with and without IHAB, respectively. (b) An angiogram instance with multiple stenoses (M-Stenosis).

TABLE III
DIFFERENT FEATURE REPRESENTATIONS (F_{HEAL}, $F_{\underline{HEAL}}$, F_{nokey}) ARE FED INTO RANDOM FOREST REGRESSOR

Method	MAE (mm) ↓	PCC (%) ↑
F_{nokey}+RF	1.4077 ± 0.6916	90.73 ± 11.81
$F_{\underline{HEAL}}$+RF	1.4898 ± 0.7585	90.15 ± 10.93
F_{HEAL}+RF	1.3831 ± 0.6967	90.34 ± 11.48

D. Effectiveness of the Keyframe-View Module

From the estimated results (Rows F_{nokey}+ RF, $F_{\underline{HEAL}}$+ RF and F_{HEAL}+ RF) shown in Table III, it can be shown that the feature representation F_{HEAL} achieves an average MAE of 1.3831 mm, versus MAE of 1.4077 mm obtained by F_{nokey} and 1.4898 mm obtained by $F_{\underline{HEAL}}$. This makes it evident that the features extracted from the keyframe view can make F_{HEAL} more expressive with respect to the stenosis indices, and therefore the lower quantification error and better correlation can be obtained. Besides, as depicted in the Fig. 7, the two-view models (Rows 5, 6, i.e., Main+Key, Sup+Key) integrating keyframe view outperform the single-view models (Rows 1, 2, i.e., Main, Sup). Because, the multi-scale features extracted from keyframe view can further improve the correlation between the stenosis feature representation and the stenosis indices.

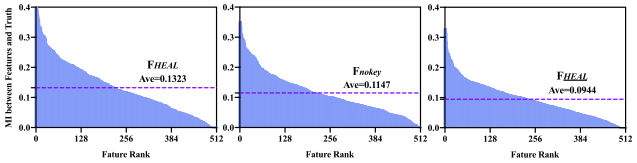

Fig. 11. Mutual information (MI) between the different feature values and the ground truth. The features are sorted by decreasing mutual information scores. The purple dashed lines indicate the average of the mutual information scores per feature.

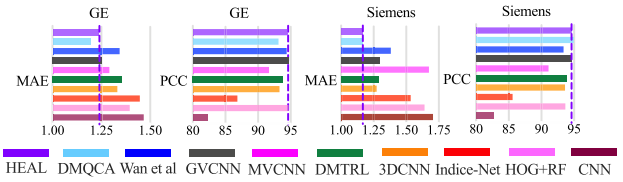

Fig. 12. Better performance of our method evaluated on the XRA data from the other two manufacturers (GE Healthcare and Siemens). The purple dashed lines show the performance of our proposed method HEAL.

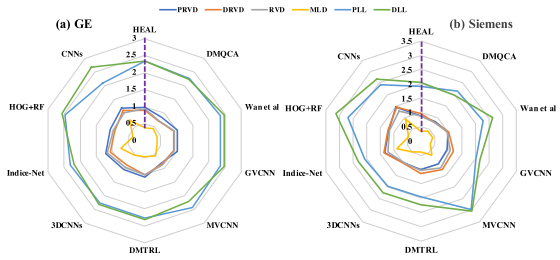

Fig. 13. Better results of our method evaluated on the XRA data from other manufacturers: (a) GE Healthcare and (b) Siemens. The purple dashed lines in radar maps indicate the results of HEAL.

In Fig.11, we also consider the information content of each individual feature of F_{HEAL}, $F_{\underline{HEAL}}$ and F_{nokey}. We measure the mutual information score between the feature value and the ground truth label. As shown in Fig. 11, the features are sorted by decreasing the mutual information scores. The MI score per feature of F_{HEAL} (0.1323) is higher than the MI score of F_{nokey} (0.1147) and that of $F_{\underline{HEAL}}$ (0.0944), thereby suggesting that F_{HEAL} contains more discriminative information related to the stenosis indices.

E. Robustness to the Difference in Manufacturers

Fig. 12 and Fig. 13 demonstrate better performance of our HEAL evaluated on the XRA data from the manufacturers compared to nine other methods. For the XRA images from GE Healthcare and Siemens, HEAL achieves lower MAE and higher PCC than the comparing methods, as shown in Fig. 12. The radar map (Fig. 13) shows the quantification error of the six clinical indices. The purple dashed lines indicate the results of our HEAL.

F. Computational Efficiency of HEAL

In terms of computational efficiency (test time) of our direct quantification method, we compute the processing time for

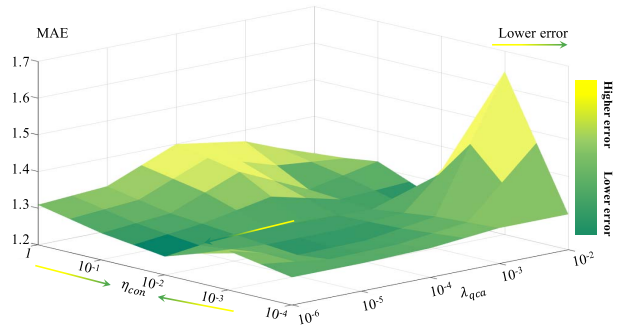

Fig. 14. Experimental results of our proposed method for stenosis quantification using different settings of two parameters.

one subject (two XRA image sequences and one keyframe, 21 frames in total). On a modern GPU (NVIDIA GTX 1080TI), our model HEAL performed very fast and used on an average only 0.06 s to process one subject. With a single CPU core (Intel Xeon E5-2620), only 0.16 s is needed. In the PCI procedure, the frame rate of all XRA sequences is commonly 15 frames/s or 7.5 frames/s. The average running time reported in the existing coronary artery diameter estimation work [10] is around 4.20 s with a CPU core. However, our method has a bigger potential to run in real-time with a more optimized implementation, to further facilitate the PCI procedure.

G. Influences of Parameters on Performance

In this section, we study the effects of two hyper-parameters, i.e., η_{con} and λ_{qca}. Specifically, we set the values in the ranges of $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$ for η_{con} and λ_{qca}, respectively. We fix the value of one parameter and tune the other parameter. Fig. 14 presents the performances of our method for stenosis quantification by using different parameter values. From Fig. 14, the experimental results demonstrate that our method HEAL can obtain the better quantification accuracy when the values of η_{con} and λ_{qca} fall in $[10^{-3}, 10^{-1}]$ and $[10^{-4}, 10^{-5}]$, respectively. Specifically, our method achieves the best performance when η_{con} and λ_{qca} are set as 10^{-2} and 10^{-6}.

H. Limitation and Future Work

Our model in the current study is limited to the direct quantification of single stenosis from XRA images. We could extend the proposed method for dealing with the direct quantification for multiple stenoses simultaneously. Secondly, our current model does not deal with the more complex stenosis lesion with other specific characteristics (e.g., diffuse, bifurcation lesion, extremely angulated segment). The intracoronary information from the other imaging modalities (e.g., from intravascular ultrasound, optical coherence tomography) could be integrated to quantify the more complex stenosis lesion. Additionally, integrating the stenosis location information for overall learning directly may not improve the quantification performance. The relationship between the two heterogeneous tasks (stenosis quantification and stenosis localization) could be learned by multi-task learning in the future.

VI. Conclusion

In this work, we have proposed a hierarchical attentive multi-view learning model HEAL for direct quantification of coronary artery stenosis from X-ray angiography (XRA) images. Our HEAL model utilizes the complementary information of stenosis in XRA sequential images from 2 views. An intra-view hierarchical attentive block is proposed to learn the discriminative feature of the stenosis. Then, a keyframe view is developed to enhance the stenosis representation by extracting multi-scale features. We have evaluated HEAL on a clinical multi-manufacturer dataset, and the experimental results show the superior quantification accuracy and better clinical agreement between the prediction and the ground truth. This endows our proposed HEAL method with a great potential for a more efficient intraoperative treatment of coronary artery disease.

References

[1] S. Tu *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," *Cardiovascular Interventions*, vol. 7, no. 7, pp. 768–777, 2014.

[2] W. Cong, J. Yang, D. Ai, Y. Chen, Y. Liu, and Y. Wang, "Quantitative analysis of deformable model-based 3-D reconstruction of coronary artery from multiple angiograms," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 2079–2090, Aug. 2015.

[3] A. J. Klein *et al.*, "Safety and efficacy of dual-axis rotational coronary angiography vs. Standard coronary angiography," *Catheterization Cardiovascular Interventions*, vol. 77, no. 6, pp. 820–827, May 2011.

[4] S. Çimen, A. Gooya, M. Grass, and A. F. Frangi, "Reconstruction of coronary arteries from X-ray angiography: A review," *Med. Image Anal.*, vol. 32, pp. 46–68, Aug. 2016.

[5] P. J. Kilner, T. Geva, H. Kaemmerer, P. T. Trindade, J. Schwitter, and G. D. Webb, "Recommendations for cardiovascular magnetic resonance in adults with congenital heart disease from the respective working groups of the European society of cardiology," *Eur. Heart J.*, vol. 31, no. 7, pp. 794–805, Apr. 2010.

[6] R. Shah *et al.*, "Comparison of visual assessment of coronary stenosis with independent quantitative coronary angiography: Findings from the prospective multicenter imaging study for evaluation of chest pain (PROMISE) trial," *Amer. Heart J.*, vol. 184, pp. 1–9, Feb. 2017.

[7] P. T. Campbell, E. Mahmud, and J. J. Marshall, "Interoperator and intraoperator (in)accuracy of stent selection based on visual estimation," *Catheterization Cardiovascular Interventions*, vol. 86, no. 7, pp. 1177–1183, Dec. 2015.

[8] A. K. Klein, F. Lee, and A. A. Amini, "Quantitative coronary angiography with deformable spline models," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 468–482, Oct. 1997.

[9] J. P. Janssen, A. Rares, J. C. Tuinenburg, G. Koning, A. J. Lansky, and J. H. C. Reiber, "New approaches for the assessment of vessel sizes in quantitative (cardio-)vascular X-ray analysis," *Int. J. Cardiovascular Imag.*, vol. 26, no. 3, pp. 259–271, Mar. 2010.

[10] T. Wan, H. Feng, C. Tong, D. Li, and Z. Qin, "Automated identification and grading of coronary artery stenoses with X-ray angiography," *Comput. Methods Programs Biomed.*, vol. 167, pp. 13–22, Dec. 2018.

[11] D. Zhang, G. Yang, S. Zhao, Y. Zhang, H. Zhang, and S. Li, "Direct quantification for coronary artery stenosis using multiview learning," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 449–457.

[12] J. Fan *et al.*, "Multichannel fully convolutional network for coronary artery segmentation in X-ray angiograms," *IEEE Access*, vol. 6, pp. 44635–44643, 2018.

[13] Y. Zhao *et al.*, "Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 438–450, Feb. 2018.

[14] S. Sankaran, M. Schaap, S. C. Hunley, J. K. Min, C. A. Taylor, and L. Grady, "HALE: Healthy area of lumen estimation for vessel stenosis quantification," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 380–387.

[15] Y. Hong *et al.*, "Deep learning-based stenosis quantification from coronary CT angiography," *Proc. SPIE*, vol. 10949, Mar. 2019, Art. no. 109492I.

[16] Z. Gao *et al.*, "Learning physical properties in complex visual scenes: An intelligent machine for perceiving blood flow dynamics from static CT angiography imaging," *Neural Netw.*, vol. 123, pp. 82–93, Mar. 2020.

[17] H. Sun, X. Zhen, C. Bailey, P. Rasoulinejad, Y. Yin, and S. Li, "Direct estimation of spinal cobb angles by structured multi-output regression," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 529–540.

[18] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2057–2067, Oct. 2017.

[19] C. Yu *et al.*, "Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 16, 2020, doi: 10.1109/TNNLS.2020.2984955.

[20] T. Zhou, K. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, Feb. 2019.

[21] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, and D. Shen, "Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis," *Sci. Rep.*, vol. 7, no. 1, pp. 1–15, Dec. 2017.

[22] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, Mar. 2002.

[23] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.

[24] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 162–190.

[25] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.

[26] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 945–953.

[27] Y. Feng, Z. Zhang, X. Zhao, R. Ji, and Y. Gao, "GVCNN: Group-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 264–272.

[28] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.

[29] H. Kim, J. Choo, J. Kim, C. K. Reddy, and H. Park, "Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2015, pp. 567–576.

[30] Z. Zhang, Z. Qin, P. Li, Q. Yang, and J. Shao, "Multi-view discriminative learning via joint non-negative matrix factorization," in *Proc. Int. Conf. Database Syst. Adv. Appl.* Cham, Switzerland: Springer, 2018, pp. 542–557.

[31] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[32] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7794–7803.

[33] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 472–480.

[34] P. Wang *et al.*, "Understanding convolution for semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2018, pp. 1451–1460.

[35] J. B. Hermiller, J. T. Cusma, L. A. Spero, D. F. Fortin, M. B. Harding, and T. M. Bashore, "Quantitative and qualitative coronary angiographic analysis: Review of methods, utility, and limitations," *Catheterization Cardiovascular Diagnosis*, vol. 25, no. 2, pp. 110–131, Feb. 1992.

[36] J. H. Reiber *et al.*, "Assessment of short-, medium-, and long-term variations in arterial dimensions from computer-assisted quantitation of coronary cineangiograms," *Circulation*, vol. 71, no. 2, pp. 280–288, Feb. 1985.

[37] C. Xu, J. Howey, P. Ohorodnyk, M. Roth, H. Zhang, and S. Li, "Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101568.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[39] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[40] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct estimation of cardiac bi-ventricular volumes with regression forests," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2014, pp. 586–593.

[41] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Med. Image Anal.*, vol. 43, pp. 54–65, Jan. 2018.

[42] A. Arbab-Zadeh and J. Hoe, "Quantification of coronary arterial stenoses by multidetector CT angiography in comparison with conventional angiography," *Cardiovascular Imag.*, vol. 4, no. 2, pp. 191–202, Feb. 2011.

[43] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, Jun. 2004, Art. no. 066138.