Information Fusion 71 (2021) 64–76

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Multi-task learning with Multi-view Weighted Fusion Attention for artery-specific calcification analysis

Weiwei Zhang [a,1], Guang Yang [b,c,1], Nan Zhang [d], Lei Xu [d,*], Xiaoqing Wang [e], Yanping Zhang [f], Heye Zhang [a,*], Javier Del Ser [g,h], Victor Hugo C. de Albuquerque [i,j]

[a] School of Biomedical Engineering, Sun Yat-sen University, 510006 Shenzhen, China
[b] Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK
[c] National Heart & Lung Institute, Imperial College London, London SW7 2AZ, UK
[d] Department of Radiology, Beijing Anzhen Hospital, Capital Medical University, 100029 Beijing, China
[e] Department of Cardiology, Fuwai Hospital, Chinese Academy of Medical Sciences Shenzhen, 518057 Shenzhen, China
[f] School of Computer Science and Technology, Anhui University, 230601 Hefei, China
[g] TECNALIA, Basque Research and Technology Alliance, 48160 Derio, Spain
[h] University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain
[i] Armtec Robotics Technology, Fortaleza, Brazil
[j] LAPISCO, Federal Institute of Education, Science and Technology of Ceará, Fortaleza, Brazil

ARTICLE INFO

Keywords:
Artery-specific calcification analysis
Multi-view learning
Multi-view Weighted Fusion Attention
Multi-task learning

ABSTRACT

In general, artery-specific calcification analysis comprises the simultaneous calcification segmentation and quantification tasks. It can help provide a thorough assessment for calcification of different coronary arteries, and further allow for an efficient and rapid diagnosis of cardiovascular diseases (CVD). However, as a high-dimensional multi-type estimation problem, artery-specific calcification analysis has not been profoundly investigated due to the intractability of obtaining discriminative feature representations. In this work, we propose a Multi-task learning network with Multi-view Weighted Fusion Attention (MMWFAnet) to solve this challenging problem. The MMWFAnet first employs a Multi-view Weighted Fusion Attention (MWFA) module to extract discriminative feature representations by enhancing the collaboration of multiple views. Specifically, MWFA weights these views to improve multi-view learning for calcification features. Based on the fusion of these multiple views, the proposed approach takes advantage of multi-task learning to obtain accurate segmentation and quantification of artery-specific calcification simultaneously. We perform experimental studies on 676 non-contrast Computed Tomography scans, achieving state-of-the-art performance in terms of multiple evaluation metrics. These compelling results evince that the proposed MMWFAnet is capable of improving the effectivity and efficiency of clinical CVD diagnosis.

1. Introduction

Artery-specific calcification analysis provides simultaneous segmentation and quantification of calcified lesions in different coronary arteries. It reveals the sites and extent of stenosis for each coronary artery in detail. Segmentation for artery-specific calcification aims to locate calcified lesions in each coronary artery. It can help capture rich calcification information, including distribution and geometric characteristics of calcified lesions. The geometric characteristic can provide high diagnostic value, i.e., larger calcified lesions may intuitively result in a higher likelihood of coronary vessel stenosis, which further cause obstructive diseases [1]. Quantification for artery-specific calcification aims to measure calcium scores of every coronary artery, including three commonly used standards: Agatston score [2], volume score [3] and mass score [4]. They directly reflect the calcified level of individual arteries, which are significantly correlated with the stenosis severity in these arteries [5]. To summarize, the artery-specific calcification analysis aims to provide detailed calcification information of each artery, in addition to whole-coronary analysis, which only estimates the overall coronary atherosclerosis burden of asymptomatic subjects.

Despite its straightforward utility for the medical expert, unfortunately, no existing work has focused on simultaneous segmentation and quantification of artery-specific calcification due to the intractability of their discriminative feature representations. In particular, segmentation

* Corresponding authors.
E-mail addresses: leixu2001@hotmail.com (L. Xu), zhangheye@mail.sysu.edu.cn (H. Zhang).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.inffus.2021.01.009
Received 8 March 2020; Received in revised form 29 November 2020; Accepted 31 January 2021
Available online 6 February 2021

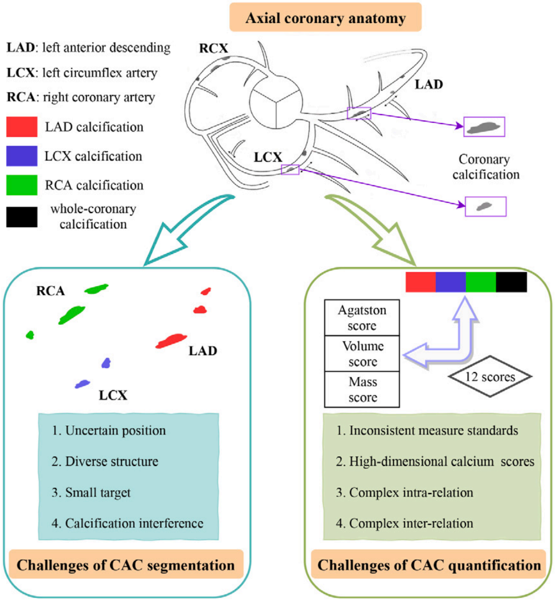

Axial coronary anatomy
LAD: left anterior descending
LCX: left circumflex artery
RCA: right coronary artery
LAD calcification
LCX calcification
RCA calcification
whole-coronary calcification
RCX
LAD
LCX
Coronary calcification
RCA
LAD
LCX
1. Uncertain position
2. Diverse structure
3. Small target
4. Calcification interference
Challenges of CAC segmentation
Agatston score
Volume score
Mass score
12 scores
1. Inconsistent measure standards
2. High-dimensional calcium scores
3. Complex intra-relation
4. Complex inter-relation
Challenges of CAC quantification

Fig. 1. The challenges of simultaneous CAC segmentation and quantification. (1) In axial coronary anatomy, the gray irregular zones represent calcified lesions with different CT intensity levels. (2) For multi-class calcification segmentation, calcified lesions have changeable shapes and uncertain location. Different colored zones represent the calcified lesions in different arteries. (3) For calcium scores estimation, multi-type calcium scores are formulated as a high-dimensional multi-output variable estimation. They have complex dependencies with calcification segmentation of corresponding arteries.

for artery-specific calcification still remains challenging because of the location and shape variability of calcified lesions. The small target characteristic of calcified lesions further increases the difficulty of this segmentation task. Moreover, rib and spine have similar pixel intensity when compared to coronary artery calcification (CAC) in Computed Tomography (CT) scan, which may result in false positives for this segmentation task. Interference also arises from the calcification in aorta and mitral valve, which not only have a similar appearance when compared to the CAC, but have also anatomically proximity to the CAC [6]. In the quantification of artery-specific calcification, different calcium scores (as the ones mentioned above) are difficult to obtain directly and simultaneously based on CT scans because of their different measure standards [2–4]. The volume score only relates to the total volume of all calcified lesions, whereas the Agatston score relies on the max CT intensity (Hounsfield units, HU) and the area of each calcified lesion [2,3]. The mass score is related to the mean CT intensity and the volume of each lesion [4]. Diverse measure standards of different scores have increased the complexity of expressive feature representations [7].

Furthermore, artery-specific calcification analysis can be modeled as a multi-task problem with complex dependencies between tasks. These dependencies not only complicate the inference of effective joint features, but also increase the intractability of joint learning. Moreover, the combination of segmentation and quantification for artery-specific calcification is also a high-dimensional multi-output problem. Mining the commonalities among multiple outputs makes it trickier in finding discriminative feature representations than single segmentation or quantification task performed in isolation. The mentioned-above challenges are illustrated in Fig. 1.

Current methods for calcium scoring follow a two-step mode, which not only lead to significant redundant work, but also neglect the multi-task correlation between the CAC location and calcium scores [6,8]. Specifically, they either detect calcified lesions and then calculate calcium scores based on the detected lesions [6], or directly estimate calcium scores, and then visualize calcified lesions based on the estimated scores [8]. Since these two-step methods try to establish an indirect mapping between CT scan and CAC, they can lead to catastrophic error propagation along the diagnostic information delivery process. Under our research hypothesis, multi-task learning may be able to solve the aforementioned problems by taking advantage of task dependency learning and commonalities mining between tasks, and further help promote the efficiency of clinical CVD diagnosis [9–12].

In this study, we validate our postulated hypothesis by proposing a Multi-task learning network with Multi-view Weighted Fusion Attention (MMWFAnet) to infer discriminative feature representations for artery-specific calcification analysis. As shown in Fig. 2, the core of MMWFAnet consists of two connected modules, i.e., Multi-view Weighted Fusion Attention (MWFA) and Multi-Task Dependency Learning (MTDL). In the CAC observation process, clinicians routinely focus on the axial view, and collect auxiliary information from coronal and sagittal views. We transfer this process as a MWFA to characterize 3D CT from these three orthogonal views [6]. Specifically, MWFA extracts expressive features from the axial view, and further exerts a weight-wise attention constraint to these features. The attention constraint is drawn from the other two views. MWFA no longer simply concatenates the features from multi-view learning, but weighs each view to more effectively combine multi-view features [13]. As a result, MWFA can help improve the discriminability of features to ease calcification analysis. Subsequently, a MTDL is proposed for simultaneous estimation of location and calcium scores of artery-specific CAC, by harnessing the fact that these two tasks have a significant intrinsic dependency. Obviously, MTDL follows the multi-task learning mechanism, which leverages information of multiple related tasks to help improve its generalization performance [9]. Unlike conventional multi-task learning, MTDL adopts a task-guided constraint between segmentation and quantification to model their task dependency. The task-guided constraint can help improve the performance of individual tasks more effectively. In general, our artery-specific calcification analysis aims to achieve estimation of the two desired objectives, i.e., segmentation and quantification of every coronary artery for a total of 16 indices, all of which are illustrated in Fig. 1. In particular, artery-specific calcification segmentation refers to the segmentation of calcified lesions in left anterior descending artery (LAD), left circumflex artery (LCX), right coronary artery (RCA) and whole-coronary for a total of 4 indices [14]. Artery-specific calcification quantification refers to the estimation of three specific calcium scores of LAD, LCX, RCA and whole-coronary for a total of 12 indices.

The main contributions of this work are summarized as follows.

- We propose a MMWFAnet for artery-specific calcification analysis. It provides multiple different calcium scores while detecting calcification in every coronary artery. Experimental results in multiple datasets show that our proposed MMWFAnet can achieve state-of-the-art performance compared to other segmentation-aware and quantification-aware modeling counterparts trained individually in terms of multiple evaluation metrics.
- We establish a Multi-view Weighted Fusion Attention (MWFA) to mimic clinical CAC observation to effectively improve the discriminability of feature representations. MWFA exploits multi-view collaboration with adaptive weights to build a spatial-channel attention model, which can enhance the effectiveness of feature representations.
- We introduce task-guided constraint into the multi-task learning to form a Multi-Task Dependency Learning (MTDL), which learns task dependencies between the segmentation and quantification tasks of artery-specific CAC. Specifically, MTDL exerts a segmentation-guided constraint on task-aware feature learning to learn the task dependency. Experimental results demonstrate that the MTDL is able to model correlation of the two tasks more effectively.

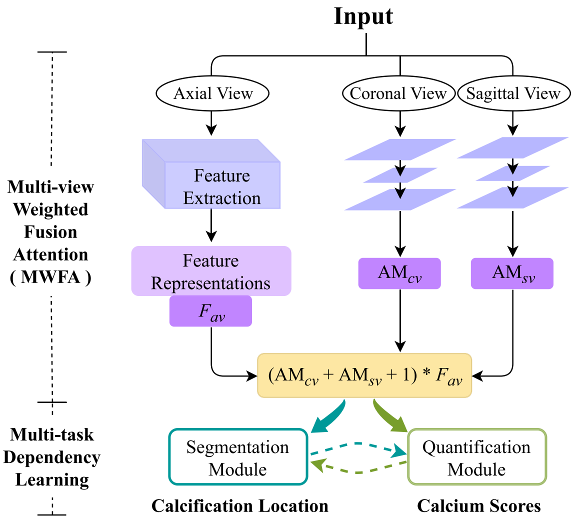


Fig. 2. Schematic diagram of the proposed MMWFAnet, which uses a MWFA to extract discriminative CAC features by multi-view information, and then realizes both segmentation and quantification of CAC via MTDL. It should be note that the coronal and sagittal views adopt the same encoder structure with different parameters. $(\mathrm{AM}_{cv}+\mathrm{AM}_{sv}+1) * F_{av}$ is consistent with Eq. (3).

The remainder of this paper is organized as follows. Section 2 reviews the related works about calcium scoring, attention model and multi-task learning. Section 3 describes in detail the proposed MMWFAnet, including the overview of our MMWFAnet, the Multi-view Weighted Fusion Attention, the Multi-Task Dependency Learning, and the optimization objective. Section 4 gives detailed descriptions of data acquisition, evaluation, and experimental settings. Section 5 reports and analyzes the experimental results. Section 6 gives the conclusions and the prospective future.

2. Related work

Based on our research purpose, we review current calcium scoring methods reported in the literature based on (1) classification of CAC, (2) regression of CAC or (3) artery-specific CAC. Most of the reviewed methods first detect calcified lesions based on classification algorithm, and then calculate calcium scores, which we call 'classification methods for CAC' (Section 2.1). A small part of the methods directly estimate calcium score based on regression algorithm, which we call 'regression methods for CAC' (Section 2.2). Regression methods for CAC exclude the detection procedure of calcified lesions in classification methods for CAC, resulting in a more direct mapping relationship. Methods for artery-specific CAC aim to realize the calcification analysis of multiple arteries simultaneously based on classification and regression methods for CAC. Subsequently, we also review available (4) attention models and (5) multi-tasking learning methods.

2.1. Classification methods for CAC

Calcified lesions are in contrast to the surrounding organs (i.e., heart and lung) in CT images, so a number of methods perform the CAC segmentation based on classification algorithm. These methods generally extract handcrafted features in earlier years, or automatic features in recent times to classify CAC and non-CAC. Subsequently, calcium scores are calculated based on the identified CAC [6,14,15]. For multimodal CT scans, the published classification methods can be further divided into three categories: (1) non-contrast cardiac CT (CT plain scan), (2) contrast enhanced cardiac CT angiography (CTA) and (3) non-contrast low-dose chest CT. CT plain scan is routinely used as the gold standard of CTA and chest CT in clinical coronary calcium scoring [16]. CTA gradually draws researchers' attention due to its low radiation dose [15–19]. Low-dose chest CT enable the identification of subjects at increased cardiovascular risk [6].

- For non-contrast cardiac CT scans, most calcification methods first obtain the region of interest (ROI), and then perform rough and fine classification based on threshold principle and machine learning algorithms. Specifically, these methods need the localization of heart or the extraction of coronary artery. However, coronary artery is unclear when compared to other tissues, such as cavities, aorta, spine and ribs. To realize CAC segmentation, these methods need to localize cavities, aorta or the whole heart in advance. Within the identified region of interest (ROI), pixels are roughly classified by thresholding to obtain the candidate CAC lesions. Subsequently, machine learning algorithms are utilized for fine classification of candidate lesions based on the handcrafted features, such as location, volume, maximal/minimal intensity, texture, etc. [14].
- For contrast enhanced cardiac CTA scans, CAC calcification methods routinely perform CAC segmentation along the pre-detected coronary artery, because the contrast agent makes the coronary artery brighter than other tissues [17].
- For non-contrast low-dose chest CT scans, CAC segmentation is usually performed in the ROI of heart. Specifically, coronary artery remains not visible, similarly to non-contrast cardiac CT scans. However, the classification methods for cardiac CT cannot be directly used for chest CT scans due to the image noise and cardiac motion artifact. Hence, CAC classification methods for chest CT scans generally and firstly use different algorithms to locate the heart as the scope for further precise CAC classification [20].

In addition, recent deep learning methods have been also widely used in CAC classification, harnessing their effective and hierarchy representation ability. These methods learn a hierarchy of discriminative features from data, replacing handcrafted ones, and conduce to superior performance [6,15,21]. In a summary, all these CAC classification methods must first perform segmentation of calcified lesions, and then computation of calcium scores, which lead to significant redundant workloads, especially in large datasets.

2.2. Regression methods for CAC

Since CT scans include adequate information of calcium scores, a few new calcium scoring methods based on CAC regression have emerged. These methods directly estimate particular calcium scores, circumventing the intermediate CAC segmentation step. For example, Cano-Espinosa et al. adopt a simple neural network structure to realize direct estimation of the Agatston score [22]. In another work, de Vos et al. employ an atlas-registration convolutional network to achieve valid slice selection and 2-D warping. The achieved image slices are then used for the regression of one specific calcium score via another convolutional network [8]. However, these CAC regression methods only obtain one particular kind of calcium score at a time, and they do not take advantage of the correlation between scores. In addition, regression methods cannot directly estimate the CAC location, and calcified lesions must be reversely visualized based on the achieved calcium score [8].

2.3. Methods for artery-specific CAC

A lot of methods have been proposed for calcium scoring, including many artery-specific ones [6,14,23–25]. These artery-specific methods estimate calcium scores of different coronary arteries based on classification [6,14] or regression [8,22] of CAC, achieving remarkable performance. For example, Wolterink et al. excellently review previous calcium scoring methods, and highlight several artery-specific methods

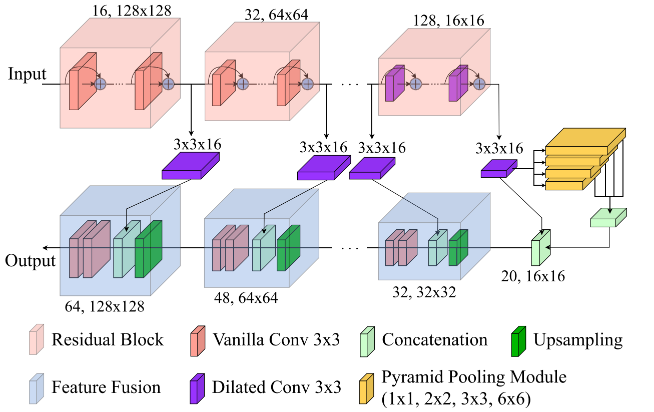


Fig. 3. Detailed structure of axial view encoder (i.e., RDN), a bottom-up and top-down architecture. In upward process, three residual blocks with vanilla convolution, one residual block with dilated convolution and a pyramid pooling module are sequentially embedded to extract multi-scale local and global CAC information. In downward process, three successive feature fusion blocks included upsampling, concatenation and vanilla convolutions are adopted to obtain CAC feature representations. The list of numbers close to residual block and feature fusion represents channels and size of output feature maps. The list of numbers close to dilated conv represents kernel size and channels. Additionally, the pyramid pooling module includes a four-level pooling of size $1\times1, 2\times2, 3\times3, 6\times6$.

based on MICCAI orCaScore challenge [24]. Hampe et al. comb machine learning-based calcium scoring methods used for calcified lesion quantification [25]. However, these methods perform segmentation and quantification of calcification step-by-step, i.e., calculating calcium scores based on the calcified lesions or visualizing calcification based on the estimated calcium scores. The step-by-step approach leads to redundant work when compared to simultaneous segmentation and quantification.

2.4. Current attention models

Attention model focuses on every input element or part of the input [26]. It has been employed to improve the learned representations of convolutional neural networks for semantic segmentation, caption generation, and natural language processing [27–29]. As the most widely used attention model, self-attention exploits input-dependent weights to linearly combine the input. Common self-attention are spatial-wise or channel-wise model based on input elements [30]. Recently, spatial-channel attention has also grasped the interest of the community, which filters the input features in both spatial-wise and channel-wise domains [27,31]. Attention models have also been applied to improve the performance of CAC segmentation [21,32]. However, previous attention models for CAC segmentation are usually implemented in the spatial-wise domain, and may not effectively and comprehensively filter CAC features.

2.5. Current multi-task learning methods

Multi-task learning (MTL) aims to leverage shared information involved in multiple tasks to help improve their general performance. MTL has been widely applied in many domains, including computer vision, natural language processing, and medical image analysis [9,13]. Ruder et al. provide an overview of MTL in deep neural networks [33]. Meyerson et al. present a high-level classification of existing deep MTL methods, exposing why MTL works well from the view of parallel ordering assumption [34]. Zhang et al. explore a learning-based approach on choosing a suitable model from a range of multi-task models for a given multi-task problem [35]. Interestingly for the scope of this study, previous MTL methods have not been used to model and exploit the dependencies between CAC segmentation and quantification tasks, which can promote them efficiently.

3. Methodology

3.1. Overview of the proposed method

We propose a MMWFAnet for effective and simultaneous segmentation and quantification of calcification in different coronary arteries. As shown in Fig. 2, our proposed MMWFAnet mainly consists of two modules: (1) a Multi-view Weighted Fusion Attention (MWFA) module with effective receptive field theory, multi-view collaboration and spatial-channel attention, and (2) a Multi-Task Dependency Learning (MTDL) module, for segmentation and quantification. MWFA can enlarge the effective receptive field (ERF) of the input to obtain expressive global and local representations [36]. It further enhances the collaboration of multiple views by constructing a spatial-channel attention. The achieved discriminative representations facilitate the learning process of the target tasks. Subsequently, the MTDL utilizes the correlation of location and calcium scores of artery-specific CAC to promote each task.

3.2. MWFA for discriminative feature representation

In clinical CAD inspection process, physicians routinely focus on planar information of 3D CT scans through visual inspection of axial slices, and further collect auxiliary information from coronal and sagittal views to support their diagnosis. We mimic this diagnostic procedure to propose a MWFA for powerful and discriminative feature representations. Its input is equivalent to the output of the down-sampling performed on 3D CT. The down-sampling consists of a 7×7 convolution with stride of 2, and a 2×2 max pooling with stride of 2, all of which are executed on axial view of 3D CT. It can help to reduce feature size and extract high-resolution global features. We slice the obtained feature maps from different views to get axial, coronal and sagittal slices. They are taken as the input of MWFA and fed into different encoders to extract expressive feature representations, as shown in Fig. 2. Then, MWFA constructs a spatial-channel attention model based on the obtained multi-view features. The attention model is committed to enhancing the collaboration of multiple imaging views. Next, we will introduce the encoders and attention model in detail.

Axial encoder. We design a residual dilated network (RDN) as the axial view encoder. Fig. 3 depicts the detailed structure of this encoder. It embraces ResNet as the backbone, and is enhanced based on the dilated convolutions and ERF theory [36,37]. In the bottom-up pathway of the RDN, we take vanilla ResNet as building blocks [37]. As the network depth increases, the receptive field size of feature maps in deep layers can be gradually expanded. However, the obtained ERF of the last residual block may not cover the heart region in original input CT scan, because ERF is usually smaller than the actual receptive field according to the ERF theory by Luo et al. [36]. Based on the understanding of dilated convolution, which can enlarge the receptive field size of feature maps, we replace all vanilla convolutions in the last residual block with dilated convolutions [38]. After ResNet is reconstructed by dilated convolutions, its ERF is able to cover the entire heart region of the CT input. As a result, RDN can extract global and high semantic features of the calcified regions, which may not be find by vanilla ResNet. In addition, a pyramid pooling module is embedded following the last residual block, aimed to extract the hierarchical global contextual CAC information in the largest receptive field [39]. In the top-down pathway, three feature fusion blocks are used to fuse global information and hierarchical multi-scale features from bottom-up pathway, towards obtaining the expressive calcification feature.

Coronal and sagittal encoders. In our network design, the coronal and sagittal views are dedicated to collect complementary information of calcification. Both auxiliary views share the same encoder structure, yet with different parameters. Since we aim to obtain desired attention maps from coronal and sagittal views, the coronal and sagittal encoders have to obtain the same size feature representations as their input through a bottom-up and top-down symmetric architecture [31]. This architecture integrates dilated convolution into U-Net, namely dilated-UNet [38,40]. In the bottom-up process of the dilated-UNet, multi-scale calcification features are gathered. These features are also utilized in top-down process through the skip connections to offset possible information loss. The skip connections enlarge the receptive field of multi-scale features through a dilated convolution at each scale, where the dilation rate is set to 2. In the top-down process, feature maps with lowest resolution are then expanded to generate dense features to inference on each pixel corresponding to the input of the dilated-UNet. In a word, dilated-UNet enlarges the receptive field of vanilla U-Net to help extract calcification feature [38,40].

Multi-view Weighted Fusion Attention. We construct a Multi-view Weighted Fusion Attention (MWFA) module to strengthen the collaboration among multiple views based attention model and multi-view learning, as shown in Fig. 2. In previous multi-view learning, CT scan is routinely projected to orthogonal axial, coronal and sagittal view [6,15,41]. Coronal and sagittal view that represented planar intensity projection images are usually used as the auxiliary feature sources [41–43]. Previous studies also show that images from multiple views all conduce to learning more useful information [44,45]. Obviously, multi-view model can utilize more information of CT scan than sing-view model. In addition, sing-view model is easy to be overfitting, which can be regularized by auxiliary view information [45]. For further feature selection based on multi-view learning, we integrate the residual attention model into it to improve the feature discriminability [31]. Specifically, let us assume F_{cv} and F_{sv} are the features of the last convolution in coronal and sagittal encoders. The subsequent Sigmoid functions resize F_{cv} and F_{sv} to [0,1], which are described as W_1 and W_2 [31]. W_1 and W_2 actually depend on F_{cv} and F_{sv}, or they are the weighted form of F_{cv} and F_{sv}. Therefore, they are automatically learned by the network. W_1 and W_2 are used as the control gates for neurons of aixal feature F_{av} similar to the Highway Network [46]. Consequently, we use them as the attention maps to rectify the axial features F_{av}:

$$W_1 \cdot F_{av} + W_2 \cdot F_{av}\,, \tag{1}$$

In order to retain axial view features by residual learning, the MWFA is grown as:

$$\begin{aligned} MWFA &= W_1 \cdot F_{av} + W_2 \cdot F_{av} + F_{av} \\ &= (W_1 + W_2 + 1) \cdot F_{av}, \end{aligned} \tag{2}$$

the residual attention learning can ensure the efficient gradient propagation and prevent its vanishing. W_1 and W_2 are in a range of [0, 1]. They are adaptively changed with F_{cv} and F_{sv} without additional constraint, such as weight sharing or normalization [31,47,48]. Then, to indicate the attention of every spatial position and channel in detail, $MWFA$ is modified as [31]:

$$MWFA(x_{i,c}) = (\, f_{cv}(x_{i,c}) + f_{sv}(x_{i,c}) + 1\,) \cdot f_{av}(x_{i,c})\,. \tag{3}$$

where indices i, c indicate the index of all spatial positions and channels, respectively, and $x_{i,c}$ denotes the feature vector of the ith spatial position and cth channel. $f_{cv}(x_{i,c})$, $f_{sv}(x_{i,c})$ denote the predicted maps of coronal and sagittal encoders, and $f_{av}(x_{i,c})$ denotes the output of axial encoder. $MWFA(x_{i,c})$ represents the feature representations learned by Multi-view Weighted Fusion Attention model.

Physicians routinely perform repeated observation on the axial view of CT images, and consult the coronal and sagittal views to aid in determining the CAC location. The proposed MWFA adopts a multi-view 2D convolution strategy to replace 3D convolution. Referred to the residual attention model [31] and as shown in Eq. (2), the MWFA is formulated as the residual learning of axial view and uses coronal and sagittal views information as attention map. It is used to enhance good features and suppress noisy ones to improve the feature discriminability of the axial view. As a spatial-channel attention model, the MWFA performs feature selection in both spatial and channel dimensions. As shown in Eq. (3), by the Sigmoid function, the MWFA performs synchronous normalization for each spatial position and each channel without any additional constraint. The resulted $MWFA(x_{i,c})$ implies the interdependencies between any two positions in feature maps and any two channels. The classical residual attention is a complete self-attention model [31]. Based on the same input as the trunk branch, it designs another mask branch to improve the feature representations of trunk branch. Compared to residual attention model [31], our MWFA makes use of auxiliary view to construct attention map for feature selection. The attention maps are weights, but also implicitly contain the other views information. The weighted fusion of multiple views involves both the view-specific and cross-view interactions [49]. This not only enriches the information sources in case of limited hardware environment, but also eases the extraction of semantic features of the CAC. The learned discriminative feature representations by the MWFA are used as shared features in the following multi-task learning module.

3.3. MTDL for segmentation and quantification of artery-specific CAC

MMWFAnet models the comprehensive analysis of artery-specific CAC as a multi-task learning problem and addresses it as such. As shown in Fig. 2, the MTDL stage of MMFWAnet consists of two task-aware modules: segmentation and quantification. Firstly, the segmentation-specific module decodes the shared feature representations for CAC segmentation. It includes an upsampling layer, a concatenation operation, a convolution layer, and another upsampling layer. The first upsampling layer directly expands the shared features. The concatenation operation fuses the feature maps from closed upsampling layer and prior down-sampling module. The other upsampling layer aims to resize feature maps to original size as the same as the input of the MMWFAnet. Secondly, the quantification-specific module extracts lesion-aware features for calcium scores regression, including maximum, mean value and area of each lesion and the number of calcified pixels. Specifically, given the shared features $\mathbf{S} \in \mathbb{R}^{H \times L \times W \times C}$, we reshape it to $\mathbb{R}^{H,(L \times W \times C)}$. After two successive fully connected layers (with size of 128 × 12), it becomes to $\mathbf{Q} \in \mathbb{R}^{H,12}$. The subsequent summation is performed on the first dimension of $\mathbf{Q}$ to obtain $\mathbb{R}^{1,12}$, which yields the ultimate quantification results. The design scheme of the quantification module is consistent with the clinical knowledge that the total calcium scores are calculated by summing scores obtained from each slice.

MTDL introduces the task dependency to realize the ambition of modeling the correlation between segmentation and quantification tasks. The shared feature representations are well disentangled in different tasks that correspond to location and scores of calcified lesions, respectively. For the two types of calcification indices, significant correlation exists between and within each type of the indices, which are referred as intra-task and inter-task dependencies.

Intra-task dependency. Regardless whether CAC segmentation or quantification is considered, a strong dependency exists among the multiple outputs within each task. (1) MMWFAnet formulates the CAC segmentation problem as a multi-class classification task, so whole-coronary and artery-specific calcification present mutual constraints, e.g., in one extreme case, if the whole-coronary calcification does not exist, the artery-specific calcification must not exist either. (2) MMWFAnet formulates the CAC quantification problem as a multivariate regression task. For the same type of calcium score, if the whole-coronary score equals zero, artery-specific scores are also set to zero. Likewise, for each artery with three types of scores, if one type score is zero, the other two types scores are both zero. The specific correspondence between different arteries and different calcium scores is implied in Fig. 1. Consequently, MMWFAnet exerts task-aware constraints on the segmentation and quantification task independently, that is:

$$R_{intra} = \sum_t \|W_t\|_2^2\,,\ \text{for } t \in \{seg,\ qua\}\,. \tag{4}$$

where seg and qua respectively denote segmentation and quantification task. R_{intra} is also known as L2 regularization, which eases the common feature selection for related outputs in each task.

Inter-task dependency. CAC segmentation and quantification module extract task-aware features based on the shared feature representations, and then obtain the calcification location and calcium scores, respectively. The two tasks aim to get different representations of calcified lesions. Obviously, calcium scores change as per the CAC segmentation results. As the number of detected CAC pixels rises, calcium scores increase accordingly. If an artery does not present calcification, the corresponding calcium score is confined to zero. To penalize the violation of the inter-task dependency, a segmentation-guided constraint is applied to the quantification results, as described by:

$$\begin{aligned}
R_{inter} &= R_{inter_1} + R_{inter_2}\,,\\
R_{inter_1} &= \frac{1}{C\times S}\sum_{c,s}\varphi(\hat{y}^{c}_{seg}=0)\,(\hat{y}^{c,s}_{qua}=0)\,,\\
R_{inter_2} &= \frac{1}{C}\sum_{c}[\varphi(E_{seg}>0)\,max(-E^{c}_{qua},0)+\\
&\qquad \varphi(E_{seg}<0)\,max(E^{c}_{qua},0)]\,,\\
E_{seg} &= \sum\hat{y}^{c_1}_{seg}-\sum\hat{y}^{c_2}_{seg}\,,\ E^{c}_{qua}=\hat{y}^{c_1,s}_{qua}-\hat{y}^{c_2,s}_{qua}\,.
\end{aligned} \tag{5}$$

where $c = 1\ldots C$ and $s = 1\ldots S$ represent the calcification category and the type of calcium score, respectively. c_1, c_2 refer to two general different categories. $\varphi(\cdot)$ denotes conditional function. E_{seg} is the bias between the number of CAC pixels in two coronary arteries, and E^{c}_{qua} is the bias between the calcium scores of two coronary arteries. $\hat{y}^{c}_{seg}$ indicates the cth calcification of the segmentation task, and $\hat{y}^{c,s}_{qua}$ indicates the sth estimated calcium score of the cth calcification in quantification task. R_{inter} effectively models the intrinsic inter-dependency between the CAC segmentation and quantification task, which ensures that the estimated calcium scores are consistent with the actual coronary calcification.

3.4. Optimization objective of MMWFAnet

The loss function of the MMWFAnet is defined as:

$$L=\omega_1 L_{seg}+\omega_2 L_{qua}+\lambda_1 R_{intra}+\lambda_2 R_{inter}\,. \tag{6}$$

where L_{seg} evaluates the overlapping degree between detected calcified regions and ground-truth regions by the generalized Dice index in the segmentation task [50]. L_{qua} evaluates the score-wise intensity differences by the advanced mean absolute errors $\sum_{c,s}^{C,S}\left|\hat{y}^{c,s}_{qua}-ln(y^{c,s}_{qua}+1)\right|$ in CAC quantification task, where $\hat{y}^{c,s}_{qua}$ and $y^{c,s}_{qua}$ represent the estimated results and ground truth of sth calcium score for category c calcification, respectively [8]. The loss term L_{qua} imposes high penalties for erroneous low quantification predictions. Superior performance is therefore enforced to lower calcium burden, which is beneficial to the evaluation of CVD risk categories. ω_1, ω_2 are the balance coefficients of different types of losses L_{seg} and L_{qua} [10]. The two scalars are assigned based on the gradient size of each dependent task as it converges. They force the two losses changing in similar scales and are facilitated to reach the balance point of convergence for both tasks. Different scalars lead to the performance changing in both tasks, as shown in Fig. 4. R_{intra}, R_{inter} indicate the regularization terms that imply the intra-task and inter-task dependency in the MMWFAnet, and λ_1, λ_2 indicate their weight coefficients, which will cause fluctuations in performance of both tasks (as shown in Fig. 5).

4. Experimental setup

4.1. Data acquisition and processing

To evaluate the performance of our MMWFAnet, in total 676 non-contrast cardiac CT scans used for routinely clinical calcium scoring are collected from two medical centers (Center 1: Beijing Anzhen Hospital, Capital Medical University and Center 2: Fuwai Hospital, Chinese Academy of Medical Sciences Shenzhen). All subjects have ethical approval from the institutional review board for a retrospective observational study. Subjects without contrast enhancement were scanned by a 256-detector CT scanner with 120 kVp tube voltage, and then the acquisitions were synchronized through ECG-triggering at 70% of the R–R interval (same type scanners at two centers). Scanning process adopts a standard calcium scoring scanning protocol. All the acquired non-contrast cardiac CT slices have an in-plane resolution of 0.3–0.5 mm and different section thickness of 2.5 mm (Center1) and 3.0 mm (Center2). Those abnormal scans were excluded by radiologists, including acquisitions with consecutive repeat scanning, fewer or more slices, high level of contrast disturbance, and significant artifacts due to metal implants.

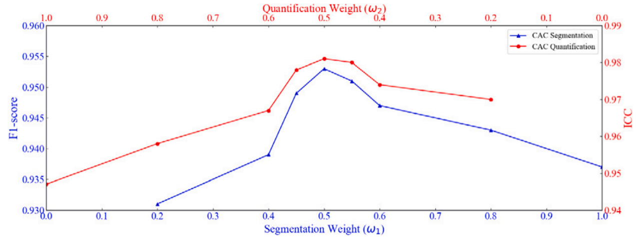


Fig. 4. Performance of both tasks as a function of the weights in loss function (ω_1, ω_2 in Eq. (6)). Since the quantification task is imposed constraints by segmentation task in our multi-task learning, its performance change is influenced by the latter.

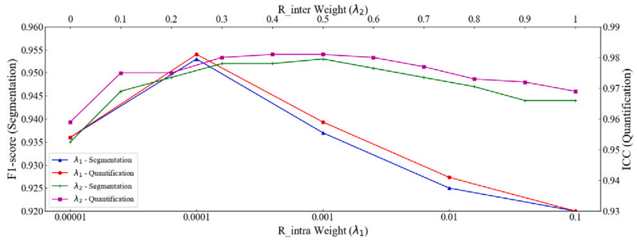


Fig. 5. Performance of both tasks as a function of the weights of normalization terms in loss function (λ_1, λ_2 in Eq. (6)). λ_1 (=0.0001) forces R_{intra} to change on the same scale with losses of both tasks. When no inter-task dependency in imposed (i.e., $\lambda_2 = 0$), their performances drop significantly.

At different medical centers, the CAC labels were provided by two experienced radiologists according to clinical calcium scoring criterion, respectively. Each radiologist independently identified and annotated the artery-specific CAC lesions, which were those connected components with a CT intensity greater than 130 HU in acquired scans. The consensus of the two radiologists for lesion annotation was used as the CAC ground truth. Subsequently, three calcium scores (e.g., Agatston score, volume score and mass score) were calculated based on the identified CAC lesions in the entire scan sequence of each subject. In addition, we adopted a strategy similar to [8] and modified it to directly split our dataset into training and testing subsets. In particular, each patient was assigned to a CVD risk category based on the calculated Agatston score (very low: <1, low: [1, 10), moderate: [10, 100), high: [100, 400), very high: ≥400) [8,14,51]. Patients in each risk category were then randomly split into training data and testing data according to a fixed ratio, eventually yielding a training dataset of 419 patients and a testing dataset of 257 patients.

4.2. Evaluation metrics and experimental settings

The performance of our MMWFAnet is comprehensively evaluated in terms of multiple types of computational metrics corresponding to different tasks [24]. (1) CAC segmentation metrics: Sensitivity, positive predictive value (PPV) and F1-score are assigned to the volume of the detected calcified lesions accordingly. They are used to measure the classification accuracy of artery-specific CAC. (2) CAC quantification metrics: Two-way intra-class correlation coefficient (ICC) for absolute agreement with 95% confidence intervals and Pearson correlation coefficient (PCC) are used to evaluate the consistency between the estimated calcium scores and the ground truth. Finally, the agreement between the estimated and ground-truth CVD risk categories is assessed by the linearly weighted Cohen's kappa coefficient κ. All the metrics vary from 0–1, with higher values implying better results.

The proposed MMWFAnet has been implemented on an Ubuntu 16.04 system with 128 GB RAM memory. It has been trained and tested using TensorFlow 1.15.0 on a NVidia Tesla P40 GPU (24 GB GPU memory). The hyperparameters are determined by model performance. We have used the Adam solver for the optimization of the network weights with an initial learning rate of 0.002 and a decay rate of 0.985 (gradually achieving a lower bound of 0.000097). The total training epochs is set to 200 and the dropout rate is set to 0.5. Algorithm indicates the optimization process of the proposed MMWFAnet.

Algorithm Multi-task learning with multi-view weighted fusion attention

```
Input: $\{x^1 \cdots x^M\}$: random minibatch of $M$ images; $\{y^1_{seg} \cdots y^M_{seg}\}$:
  corresponding calcification segmentation indices; $\{y^1_{qua} \cdots y^M_{qua}\}$: corresponding calcification quantification indices; $\omega_1$, $\omega_2$, $\lambda_1$, $\lambda_2$:
  hyperparameters;
Output: $\{\hat{y}^1_{seg} \cdots \hat{y}^M_{seg}\}$: estimated segmentation indices; $\{\hat{y}^1_{qua} \cdots \hat{y}^M_{qua}\}$:
  estimated quantification indices;
  repeat
      Compute $W_1$, $W_2$, $F_{av}$, $MWFA$ using $\{x^1 \cdots x^M\}$;
      Compute $\{y^1_{seg} \cdots y^M_{seg}\}$, $\{y^1_{qua} \cdots y^M_{qua}\}$ using $MWFA$;
      Compute $L_{seg}$, $L_{qua}$, $R_{intra}$, $R_{inter}$, $L$;
      Compute gradient $g$;
      Update $W_1$, $W_2$, $F_{av}$ using gradient $g$.
  until convergence
```

5. Results and analyses

In this section, we present the results from multiple evaluation methods, which are applied to validate the effectiveness of our MMWFAnet from different aspects. These include the comparison with the ground truth, the validation of Multi-view Weighted Fusion Attention, the validation of Multi-Task Dependency Learning and the comparison with the state-of-the-art methods.

5.1. Effectiveness of MMWFAnet compared to the ground truth

5.1.1. High accuracy compared to the ground-truth calcified lesions for the segmentation task

Tables 1 and 2 demonstrate the performance of our MMWFAnet in the CAC segmentation task. These metrics measure the pixel-wise classification accuracy of artery-specific CAC. The values of the sensitivity, PPV and F1-score for the whole-coronary calcification are 0.949, 0.960 and 0.953, respectively. These results numerically indicate that the overall coronary calcification detected by our MMWFAnet is highly overlapped with those in the ground truth.

Fig. 6 visualizes the CAC segmentation performance through the comparison between the estimation by our MMWFAnet and the ground truth for multiple arteries. It shows two example subjects that cover the scenes of multiple lesions, multiple categories, large and small lesions. As shown in Fig. 6, MMWFAnet can accurately detect all the lesions and classify them properly in CT scans with multiple lesions and multiple categories lesions. For large lesions, the segmentation results of the lesion edge by our MMWFAnet are highly consistent with those in the ground truth. Additionally, MMWFAnet is able to accurately delineate the small calcified lesions.

5.1.2. High correlation with the ground-truth calcium scores for the quantification task

Tables 1 and 2 also evince the great performance of our MMWFAnet in the CAC quantification task. Multiple calcium scores of different arteries are evaluated by ICC. The ICCs of the Agatston score, volume score and mass score for the whole-coronary calcification are 0.983, 0.983 and 0.985, respectively. These results numerically indicate that the overall coronary calcification quantified by our MMWFAnet is highly consistent with those obtained from the ground truth.

Fig. 7 shows the Bland–Altman plots for three scores of the whole-coronary calcification. The x-axis denotes the mean value of scores estimated by our MMWFAnet and the ground truth. The y-axis denotes their bias, while the mean bias is represented by the yellow–green solid line. The dark red dotted lines represent the 95% limits of agreement. Although a few outliers of high calcium scores exist that caused by the loss function (Section 3.4), which penalizes low violation to the high quantification predictions, 95.33%, 95.33% and 95.72% of the bias points are within the confidence intervals for the three calcium scores, respectively. Nevertheless, these outliers have not caused bias in the estimation of CVD risk categories (Fig. 7-d). A high Cohen's kappa score of 0.981 indicates that the estimated CVD risk categories using our MMWFAnet have a high agreement with the ones obtained from ground truth.

5.2. Effectiveness of multi-view weighted fusion attention (MWFA)

In this experiment, we evaluate the effectiveness of our Multi-view Weighted Fusion Attention through the comparisons with the single-view and the multi-view learning. The proposed MWFA exploits auxiliary views information to build an attention model, which encourages the collaboration of multiple views to facilitate the CAC segmentation. Therefore, we use axial view learning (AVL) as the baseline for comparison. Specifically, the axial encoder RDN (Section 3.2) is independently adopted as the feature selector to replace MWFA in MMWFAnet. Multi-view learning (MVL) also uses triple-encoder for three input views as the same as the MMWFAnet, but their feature representations are directly summed channel-wisely as the shared features in the following multi-task learning. The two configurations (AVL, MVL) are trained using the same strategy as the MMWFAnet, and the results in terms of multiple evaluation metrics are compared as shown in Table 2.

As summarized in Table 2, the MMWFAnet outperforms the baseline AVL and MVL in terms of multiple metrics under consideration. It numerically demonstrates the effectiveness and superiority of our MWFA relative to the RDN and the simple summation of multi-view features. We observe that the proposed MMWFAnet improves the performance in LAD calcification less obviously than other arteries when compared to the baseline AVL. As far as our knowledge, the calcified lesions in LAD routinely distribute in one or few adjacent slices, and they also have relatively large size. However, calcified lesions in LCX and RCA diffuse in multiple slices and present small target characteristics. The MWFA utilizes the auxiliary observation information about calcified lesions from other views to effectively capture small calcified lesions. It integrates multi-view information through attention model to promote the lesion-wise identification. Therefore, our MMWFAnet has a more significant effect on the calcification analysis in LCX and RCA. In addition, the AVL utilizes RDN to extract multi-scale implicit features and embeds a larger effective receptive field than vanilla ResNet, which can promote the feature representation ability of the AVL. RDN further improves the segmentation performance for inner calcified lesions. Compared to the MVL, our MMWFAnet executes further feature selection through MWFA, resulting in a superior performance than it.

Fig. 8 visualizes the obtained feature representations for LAD, LCX and RCA by our MMWFAnet, which focus on the calcified regions and have higher discriminability. In the first row, the areas pointed by pink arrows are lighter than the areas pointed by red arrows,

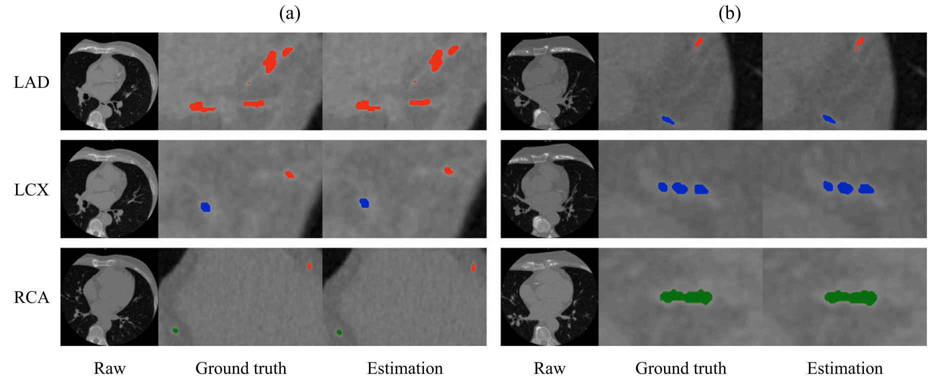


Fig. 6. Visualization of CAC segmentation performance for subject a and b. Each row represents the comparison between the estimated calcification and the ground truth of LAD, LCX and RCA, respectively. Calcified lesions routinely hold a small proportion in CT images. To display these lesions in detail and clearly show the comparison between estimation and ground truth, only the small region around calcification is highlighted (red: LAD calcification, blue: LCX calcification, green: RCA calcification). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Superior performance of MMWFAnet compared to the state-of-the-art segmentation and quantification models, including 3D FCN [52], 3D U-Net [53], 3D ResNet [37], ACNN-Seg [54], SSLLN [55], Indices-Net [56], DMTRL [57]. F1-score measures the performance of CAC segmentation based on calcified volume. ICC(AS), ICC(VS) and ICC(AS) denote the correlation between the estimated scores and the ground truth (i.e., Agatston score, volume score, mass score). "/" indicates that the F1-score metric cannot be estimated.

		3D FCN	3D U-Net	3D ResNet	ACNN-Seg	SSLLN	Indices-Net	DMTRL	MMWFAnet
F1-score	LAD	0.928	0.932	0.941	0.937	0.931	/	/	**0.961**
	LCX	0.907	0.911	0.919	0.916	0.913	/	/	**0.942**
	RCA	0.906	0.909	0.915	0.912	0.911	/	/	**0.939**
	whole	0.926	0.928	0.933	0.931	0.928	/	/	**0.953**
ICC(AS)	LAD	0.941	0.946	0.958	0.954	0.947	0.930	0.939	**0.985**
	LCX	0.929	0.937	0.943	0.940	0.936	0.921	0.932	**0.969**
	RCA	0.927	0.935	0.940	0.938	0.934	0.917	0.930	**0.969**
	whole	0.938	0.943	0.955	0.950	0.944	0.923	0.935	**0.983**
ICC(VS)	LAD	0.942	0.945	0.957	0.954	0.946	0.931	0.937	**0.986**
	LCX	0.931	0.935	0.945	0.941	0.934	0.923	0.932	**0.970**
	RCA	0.929	0.934	0.943	0.939	0.931	0.920	0.932	**0.968**
	whole	0.939	0.944	0.956	0.951	0.943	0.925	0.936	**0.983**
ICC(MS)	LAD	0.944	0.948	0.960	0.958	0.949	0.933	0.942	**0.987**
	LCX	0.933	0.939	0.948	0.944	0.938	0.925	0.935	**0.973**
	RCA	0.930	0.937	0.946	0.942	0.937	0.922	0.931	**0.972**
	whole	0.941	0.945	0.958	0.954	0.946	0.928	0.938	**0.985**

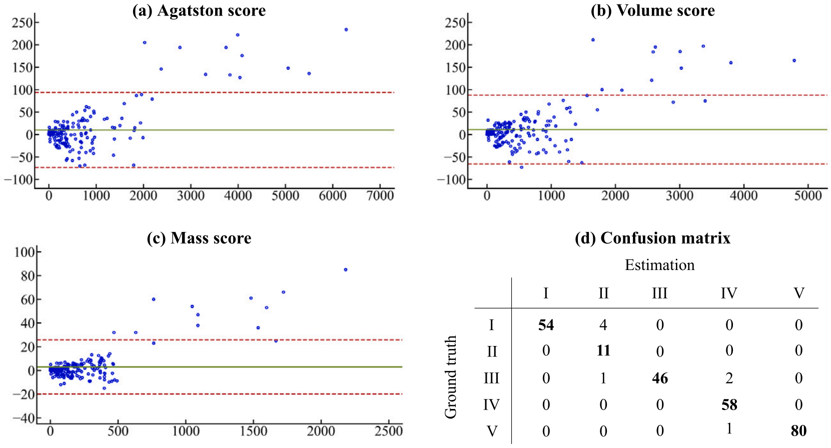
Table 2
Superior performance of MMWFAnet compared to the axial view learning (AVL), multi-view learning (MVL) and single-task learning (S-model, Q-model). *seg*, *qua* represent segmentation and quantification task. "✓" indicates that the five comparison methods contain one or more of four modules (multi-view, MWFA, *seg* or *qua*). "/" indicates that the 'Q-model' cannot acquire segmentation metrics.

			AVL	MVL	S-model	Q-model	MMWFAnet
Configuration	multi-view			✓	✓	✓	✓
	MWFA				✓	✓	✓
	seg		✓	✓	✓		✓
	qua		✓	✓		✓	✓
Segmentation	Sensitivity		0.921	0.939	0.922	/	**0.949**
	PPV		0.947	0.955	0.952	/	**0.960**
	F1-score	LAD	0.949	0.955	0.947	/	**0.961**
		LCX	0.921	0.939	0.925	/	**0.942**
		RCA	0.915	0.926	0.919	/	**0.939**
		whole	0.934	0.947	0.937	/	**0.953**
Quantification	ICC(AS)		0.966	0.975	0.960	0.947	**0.983**
	ICC(VS)		0.965	0.976	0.958	0.946	**0.983**
	ICC(MS)		0.969	0.981	0.964	0.950	**0.985**
	ICC		0.963	0.977	0.970	0.949	**0.982**
	PCC		0.969	0.985	0.980	0.961	**0.989**
	κ		0.951	0.972	0.969	0.947	**0.981**

which indicate that the MWFA can suppress these interferences while retaining objective features. In the second row, the difference indicated by different arrows demonstrates that our MMWFAnet can utilize the auxiliary views to ease the useful feature extraction. Fig. 8 intuitively indicates the effectiveness of our MWFA in CAC feature extraction.

5.3. Effectiveness of multi-task dependency learning

In this experiment, we evaluate the effective of our MTDL through the comparison with single-task learning. In Table 2, S-model and Q-model represent the learned models in independent segmentation



	I	II	III	IV	V
I	**54**	4	0	0	0
II	0	**11**	0	0	0
III	0	1	**46**	2	0
IV	0	0	0	**58**	0
V	0	0	0	1	**80**

Fig. 7. (a)(b)(c) Bland–Altman plots showing agreement between the estimated whole-coronary Agatston, volume, mass scores and the ground truth per subject, (d) Confusion matrix showing agreement in CVD risk categories based on the whole-coronary Agatston score (Five categories: I: <1, II: [1, 10), III: [10, 100), IV: [100, 400), V: ≥400). The values on the diagonal represent the number of subjects correctly estimated in each category.

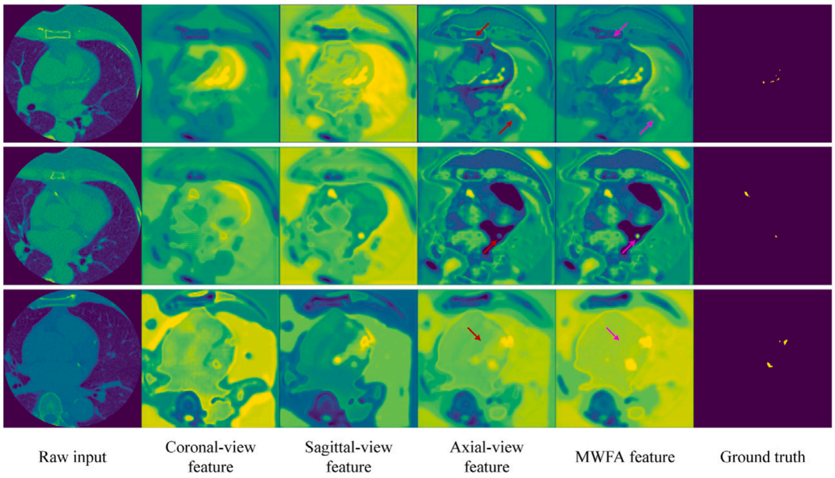


Fig. 8. Discriminative feature representations are captured by MMWFAnet through the visualization of implicit MWFA features based on Grad-CAM [58]. Because the coronal-view and sagittal-view features are converted to the axial view and then visualized, they look like axial-view feature. Coronal-view and sagittal-view features are characterized by the dilated-UNet, and axial-view feature is extracted by RDN. MWFA feature is obtained through the weighted integration of multi-view features denoted by Eq. (3).

and quantification task. S-model only realizes the artery-specific CAC segmentation, so the corresponding calcium scores, ICC and κ are calculated in accordance with the detected calcified lesions. Q-model adopts the direct estimation strategy to predict calcium scores, removing the intermediate segmentation module.

As shown in Table 2, the quantification results of Q-model are generally inferior to the calculated scores in S-model. The quantification task is struggling to learn the relationships between input images and output scores because their correspondence is not intuitive and their dimensions are inconsistent. However, independent CAC segmentation model is easier to establish the effective and intuitive mapping relationships between input image and output calcified pixels, because it aims to produce dense maps to inference on each pixel. The proposed MMWFAnet utilizes the shared MWFA features to fuse the multi-source information from segmentation and quantification labels, which enrich the auxiliary supervision signal in each task. In addition, our MMWFAnet exerts regularization constraint on the optimization process with intra- and inter-dependency of the two desired tasks, which further facilitate the use of segmentation information in the quantification task.

In order to intuitively show the multi-task relationships, we introduce the Pairwise Performance Gain (PPG) metric to evaluate the correlation strength between different tasks, denoted by $PPG = \sqrt{P_i' \cdot P_j' / (P_i \cdot P_j)}$ [59]. The P_i, P_j, P_i', P_j' represent the performance of task seg, qua when they are learned independently and with MTDL. We calculate the PPG between segmentation and quantification tasks and visualize them in Fig. 9. Color changing from white to dark blue indicates that the strength of correlation is gradually increased. As shown in Fig. 9, the CAC segmentation of an artery and corresponding multiple calcium scores possess higher PPG, which is consistent with clinical knowledge. In addition, calcification segmentation or calcium scores between branch arteries and whole coronary obtain relatively high PPG, because they have an inherent relation of part and overall. However, low PPG for scores of different branches indicate the indirect and complicated relations between them.

5.4. Comparison to state-of-the-art methods

In this section, we evaluate the effectiveness of our MMWFAnet through the comparison with the state-of-the-art methods, which include: (1) state-of-the-art segmentation and quantification methods, (2) state-of-the-art attention models and multi-task learning (MTL) models, and (3) state-of-the-art calcification analysis methods.

5.4.1. Comparison to state-of-the-art segmentation and quantification methods

Since our MMWFAnet can realize the segmentation and quantification of desired object, we compare it with the state-of-the-art models

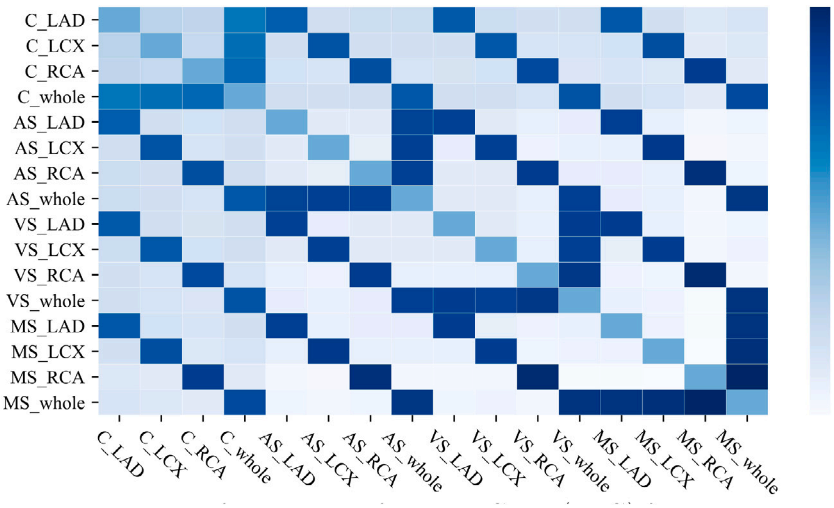


Fig. 9. The visualization of Pairwise Performance Gains (PPG) for segmentation and quantification tasks. The darker color indicates the higher PPG and the higher consistency between our estimation and the ground truth. C_LAD, C_LCX, C_RCA and C_whole represent the calcification segmentation of different coronary branches. AS, VS and MS represent Agatston score, volume score and mass score.

performed single segmentation or quantification task in isolation. The comparative models include (1) segmentation models: 3D FCN [52], 3D U-Net [53], 3D ResNet [37], ACNN-Seg [54] and SSLLN [55], and (2) quantification models: Indices-Net [56] and DMTRL [57], as listed in Table 1. Precisely because the MWFA module performs discriminative feature extraction and the MTDL module exerts intra- and inter-task dependency constraints on desired tasks, our MMWFAnet significantly outperforms the considered segmentation and quantification modeling counterparts in terms of the multiple metrics.

5.4.2. Comparison to state-of-the-art attention and MTL models

To further evaluate the performance of our MWFA module, we compare it with the state-of-the-art attention models, include Res-Att (residual attention network) [31], SENet [30] and DANet [27]. We replace the MWFA by the considered attention models in our MMWFAnet, and the comparative results are shown in Table 3. Res-Att uses simple sigmoid to realize the normalization of each spatial position and channel [31]. DANet proposes a position attention module to learn the spatial interdependencies and a channel attention module to model the channel interdependencies [27]. The two models are suboptimal than our MMWFAnet, because they do not take the weighted fusion of multi-view information into consideration. But they exceed SENet, which only and adaptively recalibrates channel-wise feature responses [30].

To further evaluate the performance of our MTDL module, we compare it with the state-of-the-art MTL models, include DMTRL [57], HU-MTL (Homoscedastic uncertainty MTL) [10] and ARM-MTL (adversarial reverse mapping MTL) [11]. These MTL models are used in our MMWFAnet, instead of the proposed MTDL. As shown in Table 3, DMTRL only realizes the quantification estimation, and cannot make use of the segmentation-based constraint [57]. HU-MTL focuses on the better weights among multiple tasks, but actually learns the implicit task dependency based on homoscedastic uncertainty [10]. The well performance of ARM-MTL originates more from its adversarial reverse mapping mechanism and bidirectional parameter sharing scheme, rather than its MTL module [11].

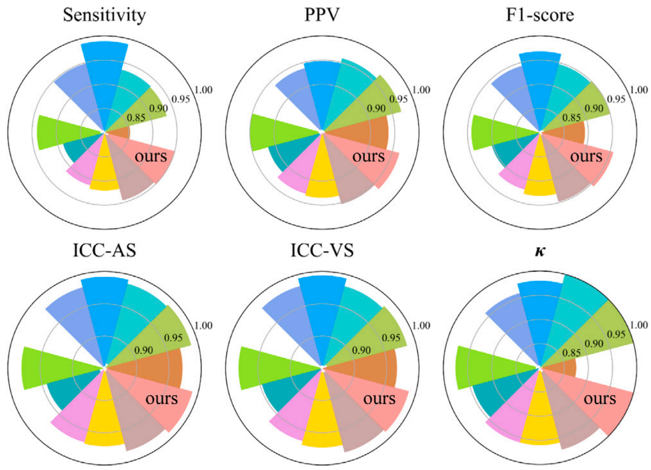

Sensitivity
PPV
F1-score
ICC-AS
ICC-VS
κ

Fig. 10. Effectiveness of MMWFAnet validated by the comparison with previous segmentation-based calcification analysis methods on orCaScore dataset. Three metrics in the first row, i.e., sensitivity, PPV, F1-score, are used to measure the CAC segmentation performance, and the other three are used to evaluate the consistency of calcium scores calculated based on the detected lesions. Starting from the chocolate bar (🟧) counterclockwise, the first six methods are from the orCaScore challenge, and the other six are successively Wolterink et al. [14] (🟩), Shahzad et al. [14] (🟦), Saur et al. [14] (🟪), Wolterink et al. [15] (🟨), Lessmann et al. [6] (🟫) and our MMWFAnet (🟥). The missing part in the upper left corner of each plot indicates that the method is inferior and not comparable. It is of note that the blue bar (🟦) with high values represents a semi-automatic method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.4.3. Comparison to state-of-the-art calcification analysis methods

In this experiment, we evaluate our MMWFAnet through the comparison with previous calcification analysis methods. Since the concept of MMWFAnet for simultaneous segmentation and quantification of artery-specific CAC is novel, no counterpart in previous methods is comparable. Therefore, we compare it with individual CAC segmentation methods and the direct CAC quantification methods based on the benchmark MICCAI orCaScore dataset with evaluation metrics [60]. The public dataset consists of 32 train patients and 40 test patients, both of them include similar non-contrast enhanced cardiac CT with our dataset. Our model performs simple fine-tuning on the training data and are evaluated on the testing data.

Fig. 10 shows the comparison of MMWFAnet with the previous CAC segmentation methods. Results of the first six methods directly come from the MICCAI orCaScore challenge, while the other comparable methods are realized by our own reimplementation based on the descriptions in the literature [6,14,15,60]. Fig. 10 indicates that our MMWFAnet achieves the state-of-the-art performance in terms of all the metrics.

Table 3
Superior performance of MMWFAnet compared to previous attention and multi-task learning (MTL) models. The comparative attention models include Res-Att (residual attention network) [31], SENet [30] and DANet [27]. The comparative MTL models include DMTRL [57], HU-MTL (Homoscedastic uncertainty MTL) [10] and ARM-MTL (adversarial reverse mapping MTL) [11]. "/" indicates that the F1-score metric cannot be estimated.

		Res-Att	SENet	DANet	DMTRL	HU-MTL	ARM-MTL	MMWFAnet
F1-score	LAD	0.951	0.944	0.952	/	0.951	0.955	**0.961**
	LCX	0.938	0.932	0.939	/	0.934	0.939	**0.942**
	RCA	0.936	0.930	0.938	/	0.931	0.938	**0.939**
	whole	0.941	0.936	0.940	/	0.938	0.942	**0.953**
ICC(AS)	LAD	0.980	0.972	0.979	0.939	0.974	0.980	**0.985**
	LCX	0.967	0.959	0.967	0.932	0.958	0.963	**0.969**
	RCA	0.965	0.957	0.966	0.930	0.958	0.961	**0.969**
	whole	0.979	0.969	0.977	0.935	0.969	0.977	**0.983**
ICC(VS)	LAD	0.981	0.973	0.979	0.937	0.975	0.981	**0.986**
	LCX	0.968	0.960	0.968	0.932	0.959	0.963	**0.970**
	RCA	0.964	0.958	0.966	0.932	0.961	0.963	**0.968**
	whole	0.977	0.969	0.976	0.936	0.970	0.978	**0.983**
ICC(MS)	LAD	0.983	0.975	0.983	0.942	0.978	0.983	**0.987**
	LCX	0.969	0.962	0.970	0.935	0.960	0.965	**0.973**
	RCA	0.967	0.961	0.968	0.931	0.963	0.966	**0.972**
	whole	0.980	0.971	0.979	0.938	0.972	0.981	**0.985**

Only a few direct CAC quantification methods have been proposed previously. Cano-Espinosa et al. [22] adopted a relatively shallow neural network to realize the whole-coronary Agatston score estimation, and then the learned inferior non-linear mapping resulted in a lower Cohen's kappa score of 0.9. Furthermore, de Vos et al. [8] combined the classical atlas-registration technology and deep neural networks to realize a multi-modal CAC quantification. The atlas-registration can be considered as a pre-processing step, which greatly improved the performance of the calcium scoring with an ICC of 0.980 and a Cohen's kappa score of 0.980. However, our MMWFAnet adopts a knowledge-based attention module to extract discriminative CAC feature, and fully utilizes the CAC location information to facilitate the estimation of calcium scores, which achieves a higher ICC of 0.991 and a Cohen's kappa score of 1.

6. Conclusion

In this paper, we propose a multi-task learning framework named MMWFAnet to enable simultaneous segmentation and quantification of calcifications in different coronary arteries. The proposed MMWFAnet adopts a new Multi-view Weighted Fusion Attention to incorporate multi-view information of CT scans for discriminative feature extraction. Then, it further combines a multi-task learning module to solve the CAC segmentation and quantification, which integrates the task dependency into multi-task learning more effectively by modeling the correlation between tasks. The compelling performance obtained from multi-source non-contrast enhanced cardiac CT scans and the comparison with the state-of-the-art models can demonstrate that the effectiveness of our MMWFAnet in clinical CVD diagnosis.

Explainability is of utmost importance in the medical domain due to its impact on the practical adoption of new technology advances by the expert performing the diagnosis, which translates to the necessity for assessing the knowledge grasped by the model [61]. The model possesses a certain explainability through the feature visualization in Fig. 6. Future research efforts will be further invested towards checking whether the focus of the model conforms to the knowledge of diagnosis experts. If divergences arise, we will elaborate on whether the trained model unveils unknown diagnostic drivers by resorting to the latest reported advances in post-hoc model explainability [61].

The generalization of the proposed multi-view attention strategy has been demonstrated through the comparison with previous attention models and the evaluation on public dataset. Furthermore, we will actively investigate whether it can be extrapolated to problems and tasks emerging from other disciplines beyond medical diagnosis.

The proposed method belongs to the wide family of supervised learning models. It relies on the accurate labeling from physicians, which demands significant time and energy. For part of calcification analysis methods (i.e., regression methods in Section 2.2), the requirements for annotated data are easily met with clinical software. However, the segmentation branch of our model needs additional lesions labeling by physicians. Consequently, we plan to explore the possibilities of semi-supervised or unsupervised learning towards alleviating these or other issues (e.g., human labeling bias). In addition, our method has only been validated over non-contrast cardiac CT. Whether it can be applied to multiple modalities (e.g., CTA and non-contrast chest CT) requires further investigation, which is also in the research agenda stemming from this work.

CRediT authorship contribution statement

Weiwei Zhang: Methodology, Software, Formal analysis, Data curation, Writing - original draft. **Guang Yang:** Methodology, Validation, Writing - original draft. **Nan Zhang:** Investigation, Formal analysis. **Lei Xu:** Conceptualization, Resources, Supervision, Funding acquisition. **Xiaoqing Wang:** Investigation, Formal analysis. **Yanping Zhang:** Formal analysis, Project administration. **Heye Zhang:** Conceptualization, Resources, Data curation, Project administration, Visualization, Funding acquisition. **Javier Del Ser:** Formal analysis, Writing - review & editing, Visualization, Supervision, Funding acquisition. **Victor Hugo C. de Albuquerque:** Writing - review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

J. Del Ser receives funding support from the Consolidated Research Group MATHMODE (IT1294-19) granted by the Department of Education of the Basque Government, as well as from the ELKARTEK and EMAITEK funding programs of the Department of Economic Development and Infrastructures of the same institution. V. H. C. de Albuquerque receives support from the Brazilian National Council for Research and Development (CNPq, Grants #304315/2017-6 and #430274/2018-1). L. Xu receives support from a grant from the National Natural Science Foundation of China (U1908211), a grant from the Capital Medical Development Research Foundation of China (PXM2020_026272_000013), and a grant from the National Key Research and Development Program of China (2016YFC1300300). H. Zhang receives funding support from the National Natural Science Foundation of China (Grants U1801265 and 61771464), as well as a grant from the Key Program for International Cooperation Projects of Guangdong Province (2018A050506031), a grant from the Key-Area Research and Development Program of Guangdong Province (2019B010110001), the Fundamental Research Funds for 73 the Central Universities (19lgzd36), and a grant from Guangdong Natural Science Funds for Distinguished Young Scholar (2019B151502031). G. Yang receives support from the European Research Council Innovative Medicines Initiative on Development of Therapeutics and Diagnostics Combatting Coronavirus Infections Award 'DRAGON: rapiD and secuRe AI imaging based diaGnosis, stratification, fOllow-up, and preparedness for coronavirus paNdemics' (H2020-JTI-IMI2 101005122), and the AI for Health Imaging Award 'CHAIMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management' (H2020-SC1-FA-DTS-2019-1 952172).

References

[1] Z. Qian, H. Anderson, I. Marvasty, K. Akram, G. Vazquez, S. Rinehart, S. Voros, Lesion-and vessel-specific coronary artery calcium scores are superior to whole-heart agatston and volume scores in the diagnosis of obstructive coronary artery disease, J. Cardiovasc. Comput. Tomogr. 4 (6) (2010) 391–399.

[2] A.S. Agatston, W.R. Janowitz, F.J. Hildner, N.R. Zusmer, M. Viamonte, R. Detrano, Quantification of coronary artery calcium using ultrafast computed tomography, J. Am. Coll. Cardiol. 15 (4) (1990) 827–832.

[3] T.Q. Callister, B. Cooil, S.P. Raya, N.J. Lippolis, D.J. Russo, P. Raggi, Coronary artery disease: Improved reproducibility of calcium scoring with an electron-beam CT volumetric method, Radiology 208 (3) (1998) 807–814.

[4] C. Hong, C. Becker, U. Schoepf, B. Ohnesorge, R. Bruening, M. Reiser, Absolute quantification of coronary artery calcium in non-enhanced and contrast enhanced multidetector-row CT studies, Radiology 223 (2) (2002) 474–480.

[5] B.D. Rosen, V. Fernandes, R.L. McClelland, J.J. Carr, R. Detrano, D.A. Bluemke, J.A. Lima, Relationship between baseline coronary calcium score and demonstration of coronary artery stenoses during follow-up: MESA (multi-Ethnic Study of Atherosclerosis), JACC: Cardiovasc. Imaging 2 (10) (2009) 1175–1183.

[6] N. Lessmann, B. van Ginneken, M. Zreik, P.A. de Jong, B.D. de Vos, M.A. Viergever, I. Išgum, Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions, IEEE Trans. Med. Imaging 37 (2) (2018) 615–625.

[7] S. Pang, Z. Su, S. Leung, I.B. Nachum, B. Chen, Q. Feng, S. Li, Direct automated quantitative measurement of spine by cascade amplifier regression network with manifold regularization, Med. Image Anal. 55 (2019) 103–115.

[8] B.D. de Vos, J.M. Wolterink, T. Leiner, P.A. de Jong, N. Lessmann, I. Išgum, Direct automatic coronary calcium scoring in cardiac and chest CT, IEEE Trans. Med. Imaging 38 (9) (2019) 2127–2138.

[9] Y. Zhang, Q. Yang, A survey on multi-task learning, 2017, arXiv preprint arXiv:1707.08114.

[10] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7482–7491.

[11] C. Yu, Z. Gao, W. Zhang, G. Yang, S. Zhao, H. Zhang, Y. Zhang, S. Li, Multitask learning for estimating multitype cardiac indices in MRI and CT based on adversarial reverse mapping, IEEE Trans. Neural Netw. Learn. Syst. (2020) 1–14.

[12] M. Li, W. Zhang, G. Yang, C. Wang, H. Zhang, H. Liu, W. Zheng, S. Li, Recurrent aggregation learning for multi-view echocardiographic sequences segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 678–686.

[13] J. Zhao, X. Xie, X. Xu, S. Sun, Multi-view learning overview: Recent progress and new challenges, Inf. Fusion 38 (2017) 43–54.

[14] J.M. Wolterink, T. Leiner, R.A. Takx, M.A. Viergever, I. Išgum, Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection, IEEE Trans. Med. Imaging 34 (9) (2015) 1867–1878.

[15] J.M. Wolterink, T. Leiner, B.D. de Vos, R.W. van Hamersvelt, M.A. Viergever, I. Išgum, Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks, Med. Image Anal. 34 (2016) 123–136.

[16] J.M. Wolterink, T. Leiner, M.A. Viergever, I. Išgum, Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 589–596.

[17] R. Shahzad, M. Schaap, T. van Walsum, S. Klien, A.C. Weustink, L.J. van Vliet, W.J. Niessen, A patient-specific coronary density estimate, in: IEEE International Symposium on Biomedical Imaging, 2010, pp. 9–12.

[18] R. Nakanishi, J.A. Delaney, W.S. Post, C. Dailing, M.J. Blaha, F. Palella, M. Witt, T.T. Brown, L.A. Kingsley, K. Osawa, et al., A novel density-volume calcium score by non-contrast CT predicts coronary plaque burden on coronary CT angiography: Results from the macs (multicenter AIDS cohort study), J. Cardiovasc. Comput. Tomogr. 14 (3) (2020) 266–271.

[19] J. Simon, L. Száraz, B. Szilveszter, A. Panajotu, Á. Jermendy, A. Bartykowszki, M. Boussoussou, B. Vattay, Z.D. Drobni, B. Merkely, et al., Calcium scoring: A personalized probability assessment predicts the need for additional or alternative testing to coronary CT angiography., Eur. Radiol. (2020).

[20] I. Isgum, M. Prokop, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Automatic coronary calcium scoring in low-dose chest computed tomography, IEEE Trans. Med. Imaging 31 (12) (2012) 2322–2334.

[21] Y. Huo, J.G. Terry, J. Wang, V. Nath, C. Bermudez, S. Bao, P. Parvathaneni, J.J. Carr, B.A. Landman, Coronary calcium detection using 3D attention identical dual deep network based on weakly supervised learning, in: Medical Imaging 2019: Image Processing, 2019, 1094917.

[22] C. Cano-Espinosa, G. González, G.R. Washko, M. Cazorla, R.S.J. Estépar, Automated agatston score computation in non-ECG gated CT scans using deep learning, in: Medical Imaging 2018: Image Processing, 2018, 105742K.

[23] R. Shahzad, T. van Walsum, M. Schaap, A. Rossi, S. Klein, A.C. Weustink, P.J. de Feyter, L.J. van Vliet, W.J. Niessen, Vessel specific coronary artery calcium scoring: An automatic system, Acad. Radiol. 20 (1) (2013) 1–9.

[24] J.M. Wolterink, T. Leiner, B.D. De Vos, J.-L. Coatrieux, B.M. Kelm, S. Kondo, R.A. Salgado, R. Shahzad, H. Shu, M. Snoeren, et al., An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orcascore framework, Med. Phys. 43 (5) (2016) 2361–2373.

[25] N. Hampe, J.M. Wolterink, S.G. Van Velzen, T. Leiner, I. Išgum, Machine learning for assessment of coronary artery disease in cardiac CT: A survey, Front. Cardiovasc. Med. 6 (2019) 172.

[26] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in: International Conference on Machine Learning, 2015, pp. 2048–2057.

[27] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.

[28] H. Tang, D. Xu, N. Sebe, Y. Wang, J.J. Corso, Y. Yan, Multi-channel attention selection GAN with cascaded semantic guidance for cross-view image translation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 2417–2426.

[29] Z. Dai, Z. Yang, Y. Yang, W.W. Cohen, J. Carbonell, Q.V. Le, R. Salakhutdinov, Transformer-XL: Attentive language models beyond a fixed-length context, 2019, arXiv preprint arXiv:1901.02860.

[30] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. 42 (8) (2019) 2011–2023.

[31] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.

[32] J. Ma, R. Zhang, Automatic calcium scoring in cardiac and chest CT using DenseRAUnet, 2019, arXiv preprint arXiv:1907.11392.

[33] S. Ruder, An overview of multi-task learning in deep neural networks, 2017, arXiv preprint arXiv:1706.05098.

[34] E. Meyerson, R. Miikkulainen, Beyond shared hierarchies: Deep multitask learning through soft layer ordering, 2017, arXiv preprint arXiv:1711.00108.

[35] Y. Zhang, Y. Wei, Q. Yang, Learning to multitask, in: Advances in Neural Information Processing Systems, 2018, pp. 5771–5782.

[36] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2016, pp. 4898–4906.

[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[38] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, 2015, arXiv preprint arXiv:1511.07122.

[39] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.

[40] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[41] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. Van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sánchez, B. van Ginneken, Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1160–1169.

[42] X. Liu, F. Hou, H. Qin, A. Hao, Multi-view multi-scale CNNs for lung nodule type classification from CT images, Pattern Recognit. 77 (2018) 262–275.

[43] S. Liang, K.-H. Thung, D. Nie, Y. Zhang, D. Shen, Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images, IEEE Trans. Med. Imaging (2020) 1.

[44] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, W. Cai, Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT, IEEE Trans. Med. Imaging 38 (4) (2018) 991–1004.

[45] X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, X. Yang, L. Li, H. Li, J. Tian, Y. Zha, Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study, Eur. J. Radiol. (2020) 109041.

[46] R.K. Srivastava, K. Greff, J. Schmidhuber, Training very deep networks, in: Advances in Neural Information Processing Systems, 2015, pp. 2377–2385.

[47] L.-C. Chen, Y. Yang, J. Wang, W. Xu, A.L. Yuille, Attention to scale: Scale-aware semantic image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3640–3649.

[48] M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu, Spatial transformer networks, in: Advances in Neural Information Processing Systems, 2015, pp. 2017–2025.

[49] A. Zadeh, P.P. Liang, N. Mazumder, S. Poria, E. Cambria, L.-P. Morency, Memory fusion network for multi-view sequential learning, 2018, arXiv preprint arXiv:1802.00927.

[50] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2010) 353–367.

[51] J.A. Rumberger, B.H. Brundage, D.J. Rader, G. Kondos, Electron beam computed tomographic coronary calcium scanning: A review and guidelines for use in asymptomatic persons, in: Mayo Clinic Proceedings, Elsevier, 1999, pp. 243–252.

[52] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[53] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.

[54] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S.A. Cook, A. De Marvao, T. Dawes, D.P. O‘Regan, et al., Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation, IEEE Trans. Med. Imaging 37 (2) (2017) 384–395.

[55] J. Duan, G. Bello, J. Schlemper, W. Bai, T.J. Dawes, C. Biffi, A. de Marvao, G. Doumou, D.P. O'Regan, D. Rueckert, Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach, IEEE Trans. Med. Imaging 38 (9) (2019) 2151–2164.

[56] W. Xue, A. Islam, M. Bhaduri, S. Li, Direct multitype cardiac indices estimation via joint representation and regression learning, IEEE Trans. Med. Imaging 36 (10) (2017) 2057–2067.

[57] W. Xue, G. Brahm, S. Pandey, S. Leung, S. Li, Full left ventricle quantification via deep multitask relationships learning, Med. Image Anal. 43 (2018) 54–65.

[58] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[59] H. Zhang, L. Xiao, Y. Wang, Y. Jin, A generalized recurrent neural architecture for text classification with multi-task learning, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 3385–3391.

[60] MICCAI challenge on automatic coronary calcium scoring, 2014, https://orcascore.grand-challenge.org/.

[61] A.B. Arrieta, N. Diaz-Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, F. Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.