Combating Coronary Calcium Scoring Bias for Non-gated CT by Semantic Learning on Gated CT

Jiajian Li
CVTE Research, China
lijiajian@cvte.com

Anwei Li
CVTE Research, China
lianwei@cvte.com

Jiansheng Fang
CVTE Research, China
11949039@mail.sustech.edu.cn

Yonghe Hou, Chao Song, Huifang Yang
Yibicom Health Management Center, CVTE, China
houyonghe@cvte.com, songchao@cvte.com, yanghuifang@cvte.com

Jingwen Wang
CVTE Research, China
wangjingwen7003@cvte.com

Hongbo Liu *
CVTE Research, China
liuhongbo@cvte.com

Jiang Liu †
Southern University of Science and Technology, China
liuj@sustech.edu.cn

Abstract

Coronary calcium scoring (CCS) can be quantified on non-gated or gated computed tomography (CT) for screening cardiovascular disease (CVD). And non-gated CT is used for routine coronary artery calcium (CAC) screening due to its affordability. However, artifacts of non-gated CT imaging, pose a significant challenge for automatic scoring. To combat the scoring bias caused by artifacts, we develop a novel semantic-prompt scoring siamese (SPSS) network for automatic CCS of non-gated CT. In SPSS, we establish a sharing network with regression supervised learning and semantic supervised learning. We train the SPSS by mixing non-gated CT without CAC mask and gated CT with CAC mask. In regression supervised learning, the network is trained to predict the CCS of non-gated CT. To combat the influence of motion artifacts, we introduce semantic supervised learning. We utilize gated CT to train the network to learn more accurate CAC semantic features. By integrating regression supervised learning and semantic supervised learning, the semantic information can prompt the regression supervised learning to accurately predict the CCS of non-gated CT. By conducting extensive experiments on publicly available dataset, we prove that the SPSS can alleviate the potential scoring bias introduced by pixel-wise artifact labels. Moreover, our experimental results show that the SPSS establishes state-of-the-art performance.

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of death worldwide, posing a significant health burden [20, 31]. Early detection and treatment of individuals at high risk for CVD is crucial for effective prevention and management. Among various risk factors, coronary artery calcium (CAC) has emerged as a strong and independent predictor of cardiovascular events, including myocardial infarction [4, 18].

Quantification of CAC, i.e. coronary calcium scoring (CCS), involves quantifying the amount of calcified plaque in the coronary arteries using computed tomography (CT) scans [7, 8]. The evaluation is typically performed on dedicated non-contrast-enhanced cardiac CT scans. Calcified plaques are identified based on their density and size, and a score is assigned using established scoring algorithms such as the Agatston score [1]. The score takes into account the area, density, and location of the calcified lesions, providing an indication of the overall burden of coronary artery calcification. Higher CAC scores are associated with an increased risk of cardiovascular events, including myocardial infarction and stroke [12].

Clinically, gated CT can accurately capture static images of the heart during specific cardiac cycles, ensuring high image quality and reduced motion artifacts [17]. It is considered the gold standard for CCS, as shown in Figure 1 (a). However, gated CT often requires significant resources, which may not be operationally feasible at small centers. These include capital intensive CT machines and specialized monitoring (e.g. electrocardiogram gating and potential administration of beta-blockers). While
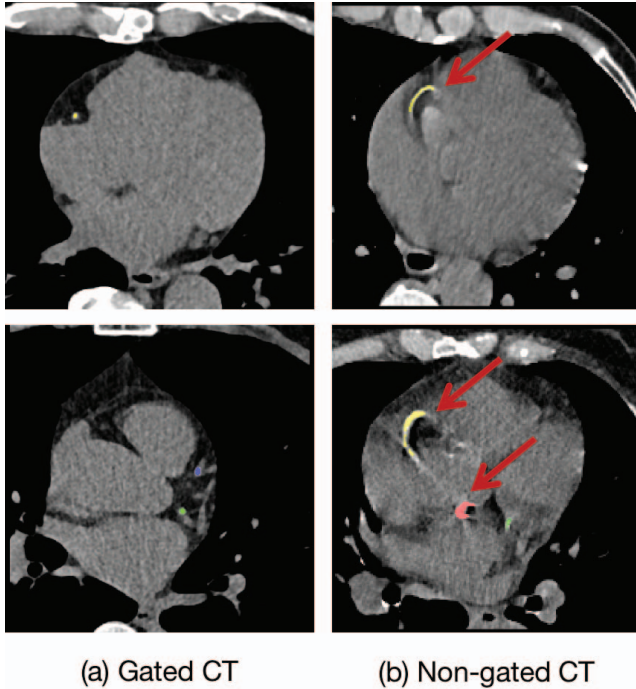
(a) Gated CT   (b) Non-gated CT

Figure 1. Examples of CAC on gated CT and non-gated CT images. Arrows indicate CAC with motion artifacts. Colors red, green, blue and yellow represent left main artery, left circumflex, left anterior descending and right coronary artery respectively.

the CCS has been traditionally measured on specialized electrocardiography-gated cardiac CT, it can also be measured on standard CT scan of the chest performed without electrocardiography (e.g. non-gated CT) [8, 26, 11]. Compare with gated CT, the non-gated CT without increasing the patient's radiation dose or economic burden [23, 22, 27]. Therefore, the quantification of CAC on non-gated chest CT becomes the first choice for routine screening.

Several semi-automatic or automatic methods have been introduced for CCS on non-gated CT, threshold-based methods [1, 9] and to the better performing deep learning approaches [15, 28, 25, 5, 3]. The threshold-based methods involve manually identifying potential regions and applying thresholding and connected component labeling to segment calcified lesions on non-gated CT. However, using a threshold alone may lead to inaccurate CAC estimation and patient risk categorization. Deep learning methods first label the CAC regions on non-gated CT scans, then use convolutional neural networks for feature extraction, and predict the CAC region and CCS through a decoder or regression model. These methods improve the accuracy and efficiency of assessing CAC, contributing to early diagnosis and treatment of coronary artery disease. However, these methods rely on non-gated CT with CAC mask. Due to the lack of electrocardiogram synchronization on non-gated CT scans, it is difficult to accurately label the true CAC region without

motion artifacts, as shown in Figure 1 (b). Obviously, these methods are difficult to combat the scoring bias caused by artifacts, which will lead to poor scoring bias. Therefore, how to reduce the scoring bias between non-gated CT and the gold standard (gated CT) has become an urgent problem that needs to be addressed.

In clinical practice, gated CT is often regarded as the gold standard due to its high-quality imaging without motion artifacts. Leveraging the semantic calcium information from gated CT to prompt deep learning models to learn more accurate calcification features on non-gated CT becomes a potential feasible method. In this work, we developed a new semantic-prompt scoring siamese (SPSS) network for CCS of non-gated CT. In SPSS, we build a shared network with regression supervised learning and semantic supervised learning. To train the SPSS network, we utilize a combination of non-gated CT without CAC masks and gated CT with CAC masks. The SPSS is trained using regression supervised learning to predict the CCS of the non-gated CT. To combat the influence of motion artifacts, gated CT scans are used to train the network to learn more accurate semantic features related to CAC, left main artery (LM), left circumflex (LCX), left anterior descending (LAD) and right coronary artery (RCA). The integration of regression supervised learning and semantic supervised learning allows the network to accurately predict the CCS of the non-gated CT by the prompt of semantic information, thus freeing from the plague of pixel-wise artifact labels.

We conduct experiments on a public Coronary Calcium and chest CT dataset to demonstrate the effectiveness of our proposed approach in different situations. Our contributions can be summarised as follow:

- We develop a novel semantic-prompt scoring siamese (SPSS) network for CCS of non-gated CT. We cleverly utilize the semantic calcium information from gated CT to combat scoring bias on non-gated CT.

- This is the first time that semantic supervised learning and regression supervised learning have been combined on non-gated CT without CAC mask.

- By conducting extensive experiments on publicly available dataset, we demonstrate the superiority of our SPSS compared to its counterparts and establish a state-of-the-art performance.

2. Related Work

Threshold-based methods. CAC is commonly defined as two or more connected voxels above 130 Hounsfield Units (HU) on gated CT [1, 19]. Threshold-based methods are semi-automatic image segmentation methods. These methods include identifying candidate regions, followed by
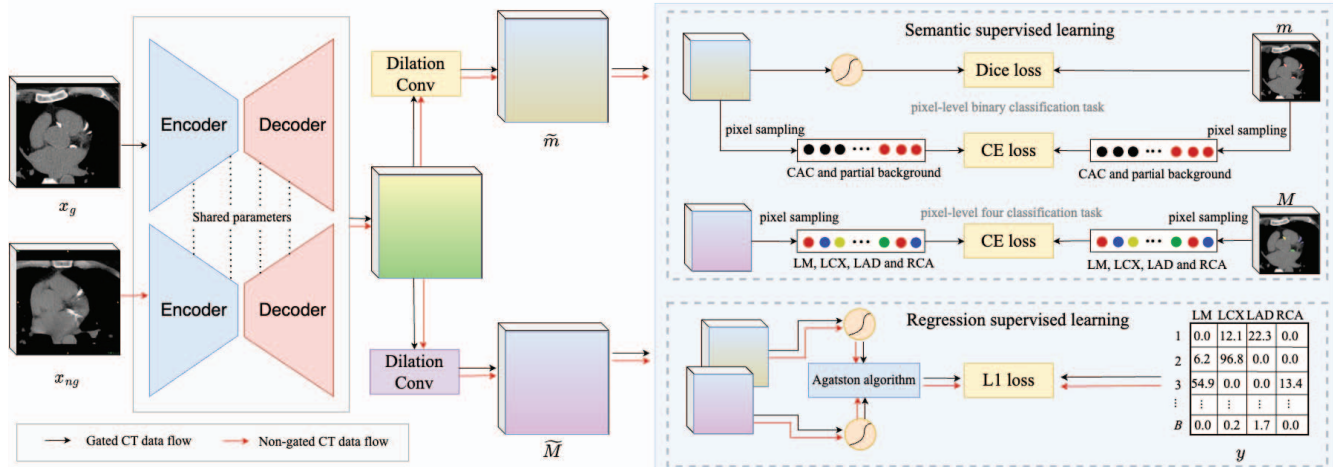
Figure 2. The architecture of our SPSS with semantic supervised learning and regression supervised learning. The black arrows indicate the data flow of 3D gated CT, and the red arrows indicate the data flow of 3D non-gated CT.

thresholding and connected component labeling. Subsequently, when lesions are identified, region growing is used to fully segment the calcified lesions. Finally, after all CAC lesions have been segmented, CAC is quantified using the Agatston score. For example, Wang et al. [29] transferred the collected scans from the CT equipment to a workstation (AW 15.0, General Electric, Boston, MA, USA) to determine the CCS using the dedicated postprocessing software "Smartscore". Thijmen et al. [9] utilized a region-growing algorithm for CAC segmentation by starting from seed points and iteratively merging neighboring pixels with pixel values greater than or equal to 130 HU. Their study demonstrated the effectiveness of the region-growing algorithm in accurately delineating CAC regions.

However, a threshold of 130 HU is not necessarily suitable for non-gated CT due to differences in equipment and imaging parameters. The use of a threshold lack robustness against noise and artifacts and may lead to under- or over-estimation of the amount of CAC on non-gated CT, leading to incorrect cardiovascular categorization of patients.

Semantic supervised learning. Most recently proposed methods employ semantic supervised learning for automatic CCS on non-gated CT [24, 15, 28, 25, 6, 5, 3]. GianmarcoIn et al. [24] trained a convolutional neural network (CNN) model using annotated CAC mask as ground truth and achieved promising results in accurately delineating CAC boundaries. In similar, Wang et al. [15] proposed a cascaded CNN architecture for CAC segmentation on non-gated CT, where the first stage learned semantic features from CT images to classify CAC presence and the second stage refined the segmentation boundaries. Their approach demonstrated improved accuracy and robustness compared to traditional segmentation methods. To more accurately identify per-artery, Bernhard et al. [6] exploited dual de-

coders to identify CAC and per-artery separately. Further, Eng et al. [5] utilized 2D SE-ResNeXt with an encoder and a decoder to segment CAC in per-artery and then quantified the predicted CAC lesions.

These approaches follow a workflow similar to current clinical calcium scoring: CAC is first identified and then quantified. Therefore, this is currently the most widely used method. However, semantic supervised methods rely on non-gated CT with CAC mask, accurately labeling the true CAC region without motion artifacts is challenging in the absence of electrocardiogram synchronization. Consequently, these methods may struggle to address scoring bias caused by motion artifacts, resulting in suboptimal quantification accuracy.

Regression supervised learning. Regression supervised learning refers to directly predicting the calcification score rather than the CAC regions on non-gated CT, and performing regression supervised learning according to the real CCS from gated CT. Direct quantification has proven to be useful for atrial and ventricle volume quantification [10, 32, 30]. Furthermore, attempts are being made to use it for CCS [2, 3]. Carlos et al. [2] used a 3D regression ConvNet for direct CCS in downsampled CT volumes also cropped around the heart. Bob et al. [3] exploited two ConvNet to align the fields of view and regress the CCS directly.

Although ConvNet cannot provide accurate CAC regions, it avoids the time-consuming CAC segmentation process and achieves similar performance compared to state-of-the-art methods. Furthermore, ConvNet does not require manual annotation of CAC regions on non-gated CT, making it suitable for addressing the interference of motion artifacts commonly present on non-gated CT.

3. Method

3.1. Problem Formulation

Given a non-gated CT training cohort $X_{ng} = \{(x_{ng_i}, y_i)\}_{i=1}^n$ and a gated CT training cohort $X_g = \{(x_{g_i}, m_i, M_i, y_i)\}_{i=1}^n$, we adopt a double (x_{ng}, y) and a quadruple (x_g, m, M, y) to stand for CAC data, including non-gated CT image x_{ng}, gated CT image x_g, binary CAC mask m, per-artery calcium mask M and coronary calcium scoring y.

Based on the gated CT training cohort with supervised information of the mask and coronary calcium score, and the non-gated CT training cohort without supervised information of the mask, we utilize regression supervised learning and semantic supervised learning to train SPSS network $f_\theta(x)$ parameterized by weights θ. The output of SPSS network includes a binary CAC map $\widetilde{m}$ and a per-artery calcium map $\widetilde{M}$, with the same size as the input. For $f_\theta(x_{ng})$ and $f_\theta(x_g)$, the parameters θ are shared in x_{ng} and x_g, as shown in Figure 2.

3.2. Pre-processing

In the pre-processing stage, we use a publicly available dataset [13] with heart annotations to train a heart localization network. This step was necessary as CT scans could differ, for example, in size, resolution, area captured, or field of view (FOV), depending on the cohort, scanner used, and site acquiring the scan. The model used for training was a standard 3D UNet [21] with four downsampling steps running for 200 epochs. After obtaining the heart position predicted by the model, we extended the predicted heart bounding box by 16 mm in the axial plane to ensure that the entire heart region is included. Next, according to the heart bounding box, all CT images are cropped to obtain a small FOV image.

3.3. Siamese Network

In SPSS, we build a shared network with semantic supervised learning and regression supervised learning for 3D gated and non-gated CT. As shown in Figure 2, we first apply an encoder-decoder structure as the backbone. Then, a small atrous convolutional block was used to capture CAC features that tend to be more local and detailed. and output the background and CAC feature map $\widetilde{m} = [\widetilde{m}_b, \widetilde{m}_c]$. Further, we use a larger atrous convolution block to capture the global features of CAC to distinguish per-artery, and output per-arterial feature map $\widetilde{M} = [\widetilde{M}_1, \widetilde{M}_2, \widetilde{M}_3, \widetilde{M}_4]$. Next, we use two softmax functions and the Agaston algorithm [1] to generate per-arterial CCS $\widetilde{y} = [\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3, \widetilde{y}_4]$. After generating predicted $\widetilde{y}$, we further utilize semantic supervised learning and regression supervised learning to train the net-

work. The k-th arterial $\widetilde{y}_k$ is defined as follows:

$$\widetilde{y}_k = Sum(\sigma(\widetilde{m})_c \cdot \sigma(\widetilde{M})_k \cdot W) \cdot s, \tag{1}$$

where s is the pixel spacing within and between slices, symbol σ denotes the softmax function. The weighted intensity W is based on the maximum radio-density in HU of a 3D lesion in the following manner: 1 = [130, 200), 2 = [200, 300), 3 = [300, 400), and 4 = [400, +∞).

3.4. Semantic Supervised Learning

Semantic supervised learning mainly includes a pixel-level binary classification task (CAC and background) and per-artery classification task (LM, LCX, LAD and RCA). To reduce the effects of background pixels on the loss function, Dice loss $\mathcal{L}_{Dice}$ and local-sampled loss $\mathcal{L}_{ls}$ are employed to train binary classification task. The $\mathcal{L}_{ls}$ is defined as follows:

$$\mathcal{L}_{ls} = -\frac{1}{N}\sum_{i=1}^N [m_i \cdot log(\widetilde{m_i}) + \mathbb{I}_i \cdot (1 - m_i) \cdot log(1 - \widetilde{m_i})], \tag{2}$$

where N denotes the pixel number, m and $\widetilde{m}$ are the binary ground truth and predicted CAC, respectively. $\mathbb{I}_i$ denotes the sampling indicator function. We define that pixel i is sampled if $\mathbb{I}_i = 1$, otherwise $\mathbb{I}_i = 0$. For the per-artery classification task, we only consider the loss of the CAC area. The loss function is defined as follows:

$$\mathcal{L}_f = -\frac{1}{N_{m>0}}\sum_{i:m>0}^{N_{m>0}}\sum_{k=1}^K M_{ik} \cdot log(\widetilde{M_{ik}}), \tag{3}$$

where M and $\widetilde{M}$ are the per-artery calcium ground truth and predicted per-artery calcium map, and K is the number of coronary arteries. The total loss of semantic supervised learning is defined as follows:

$$\mathcal{L}_{ssl} = \begin{cases} \mathcal{L}_{Dice} + \alpha\mathcal{L}_{ls} + \beta\mathcal{L}_f, & x_g \\ 0, & x_{ng} \end{cases}, \tag{4}$$

where α and β denote the weights of $\mathcal{L}_{ls}$ and $\mathcal{L}_f$, respectively. The default values of α and β both are 0.1. It is worth noting that when the training image is non-gated CT, $\mathcal{L}_{ssl} = 0$.

3.5. Regression Supervised Learning

Using regression learning to predict CCS does not require labeling of CAC regions on non-gated CT. The loss function for regression supervised learning is defined as follows:

$$\mathcal{L}_{rsl} = \frac{1}{K}\sum_{k=1}^K |y_k - \widetilde{y}_k|, \tag{5}$$

where y is the reference CCS. In addition, due to the lack of CAC reference mask on non-gated CT, it is difficult for
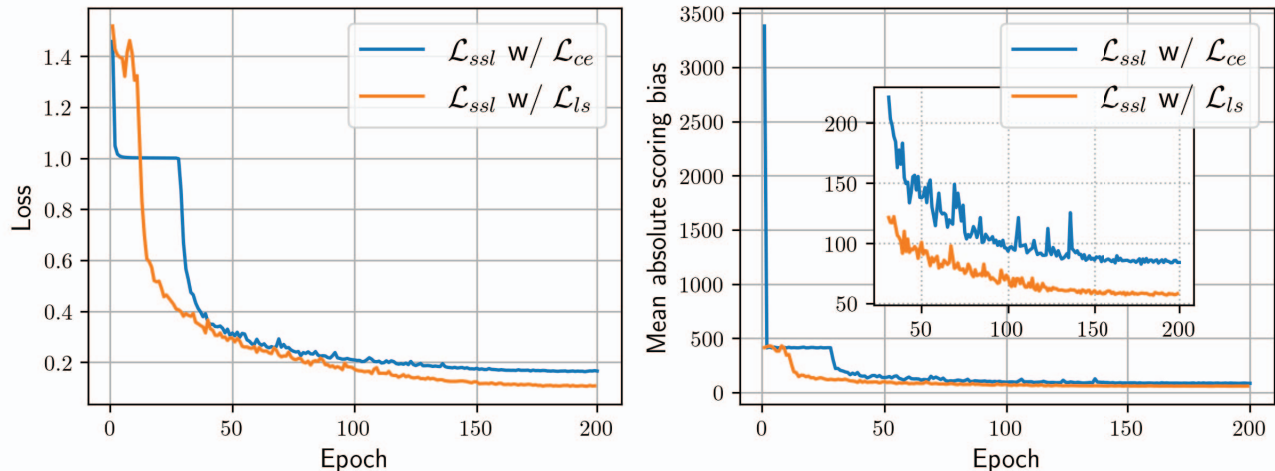
Figure 3. Convergence curves of SPSS for different loss functions on the COCA dataset. The $\mathcal{L}_{ce}$ denotes the cross-entropy loss.

Table 1. Mean error (ME) and standard deviation (SD)for per-arterial CCS on COCA test set.

Image types	Methods	Per-arterial ME $\pm$ SD				
		LM	LCX	LAD	RCA	Total
Gated CT	ConvNet	-12.81 $\pm$ 23.10	-18.10 $\pm$ 32.71	-13.30 $\pm$ 28.27	-15.15 $\pm$ 24.12	-19.36 $\pm$ 29.45
	SE-ResNeXt	-3.10 $\pm$ /	-1.99 $\pm$ /	-4.57 $\pm$ /	-2.37 $\pm$ /	-2.86 $\pm$ /
	SPSS w/o SSL	8.16 $\pm$ 20.19	1.09 $\pm$ 13.28	-14.38 $\pm$ 27.87	**1.00** $\pm$ 12.12	-4.13 $\pm$ 15.28
	SPSS	**-2.62 $\pm$ 9.21**	**0.54 $\pm$ 4.62**	**-0.11 $\pm$ 5.18**	4.49 $\pm$ 10.34	**2.31 $\pm$ 8.90**
Non-gated CT	ConvNet	-6.28 $\pm$23.47	14.77 $\pm$ 76.79	**4.39 $\pm$ 43.05**	9.90 $\pm$ 52.55	23.50 $\pm$ 68.42
	SE-ResNeXt	/	/	/	/	/
	SPSS w/o SSL	6.67 $\pm$ 45.38	**1.48** $\pm$ 27.12	-25.39 $\pm$ 64.89	-10.03 $\pm$ 54.35	-26.54 $\pm$ 62.64
	SPSS	**-0.25 $\pm$ 20.44**	-2.02 $\pm$ **21.10**	-12.30 $\pm$ 68.01	**-1.99 $\pm$ 33.40**	**-15.83 $\pm$ 44.57**

regression supervised learning to learn the semantic information of CAC. Therefore, we leverage semantic learning of gated CT to prompt regression learning. The total loss is defined as follows:

$$\mathcal{L} = \mathcal{L}_{ssl} + \lambda\mathcal{L}_{rsl}, \quad (6)$$

where λ denotes the weights of $\mathcal{L}_{rsl}$. To ensure the stability of the segmentation network, the default value of λ is 0.001.

4. Experiments

4.1. Implementation Details

Dataset. The public Coronary Calcium and chest CT's (COCA) dataset [5] from Stanford Hospital is employed in this study, which contains 447 gated CT exams in 447 patients and 212 non-gated CT exams in 212 patients. Patients who underwent non-gated CT also underwent paired gated CT exams as a reference standard for CCS. We split the patients in the COCA dataset as a training set with 507 patients (149 non-gated CT and 358 gated CT), and a test set with 152 patients (63 non-gated CT and 89 gated CT).

Methods. We experiment with four comparable methods to demonstrate the gains of our SPSS, including semantic supervised method SE-ResNeXt (Eng et al. [5]), regression supervised method ConvNet (Bob et al. [3]), SPSS without semantic supervised learning (SPSS w/o SSL) and our SPSS. The result of ConvNet is trained from scratch, and the result of SE-ResNeXt is obtained from the original paper (Eng et al. trained SE-ResNeXt by manually annotating the CAC mask on non-gated CT, but these CAC masks are not publicly available).

Evaluation metrics. Automatically predicted CCS were compared with reference CCS. Evaluations were performed on the hold-out test sets which were not used during method development. Mean error (ME) was computed to evaluate the bias between predicted and reference CCS. In addition, the agreement between predicted and reference CVD ranks was determined using Cohen's Kappa statistic (Kappa) [14].

Optimizer. Our SPSS model is trained from scratch using the SGD optimizer. Specifically, the learning rate is decay scheduled by cosine annealing [16] setting 2e-3 as an initial value and 1e-3 from 5th epochs. We set the batch
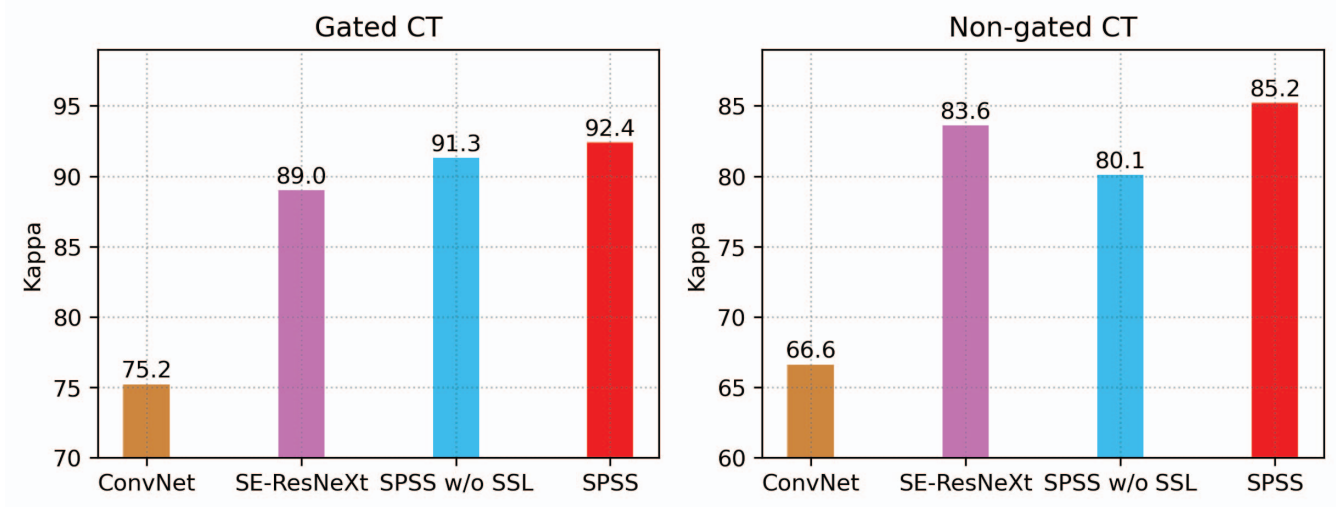
Figure 4. Histogram of CVD rank classification agreement for the four methods on gated CT and non-gated CT.
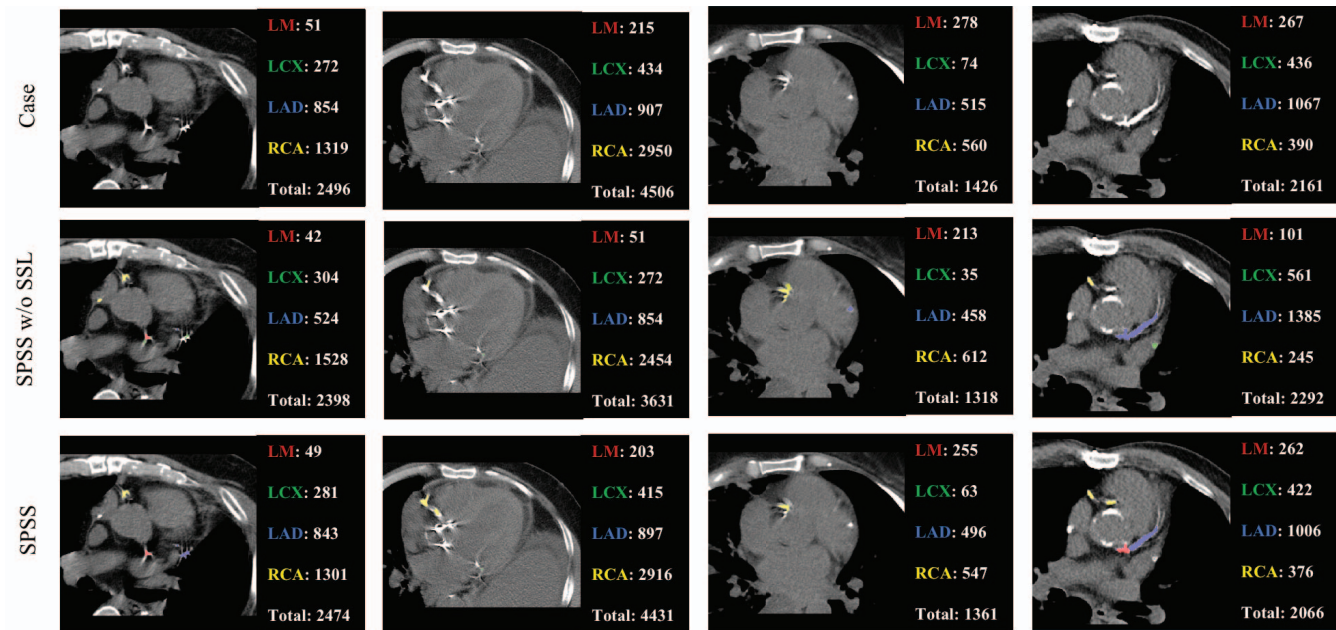

Figure 5. Examples of CAC segmentation and CCS prediction. Columns 1 to 4 are different patients, the first row is the reference image and CCS, the second row is the result of SPSS w/o SSL, and the third row is the result of our SPSS. LM, LCX, LAD and RCA are represented in red, green, blue and yellow, respectively.

size as a multiple of 2 for all tasks, i.e., 4 (2 GPUs with 2 input images per GPU). The parameters of networks are optimized in 200 epochs with a weight decay of 3e-5 and a momentum of 0.8.

4.2. Results and Analysis

Semantic learning efficiency analysis. Since the proportion of CAC in the image is very small, we must analyze the training efficiency of semantic supervised learning.

We replace $\mathcal{L}_{ls}$ with standard cross entropy $\mathcal{L}_{ce}$ in $\mathcal{L}_{ssl}$ to observe the impact of the number of negative samples on training efficiency. As shown in Figure 3, the loss and mean absolute scoring bias of $\mathcal{L}_{ssl}$ w/$\mathcal{L}_{ce}$ stagnate between the 5th and 25th epoch. This is caused by a large number of negative samples participating in gradient feedback and the contribution of positive samples being weakened. On the contrary, $\mathcal{L}_{ssl}$ w/$\mathcal{L}_{ls}$ can reduce the loss faster and reach the minimum by locally sampling the negative samples in
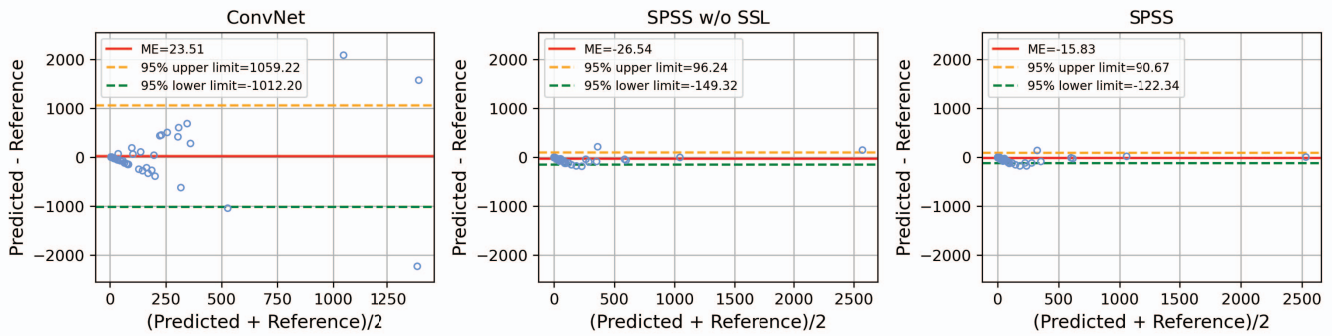
Figure 6. Bland-Altman plots showing agreement between predicted and reference per subject CCS in the non-gated CT test set. The limits of agreement are ±1.96 SD, the positive bias in ConvNet is mainly caused by overestimations of the higher CCS.
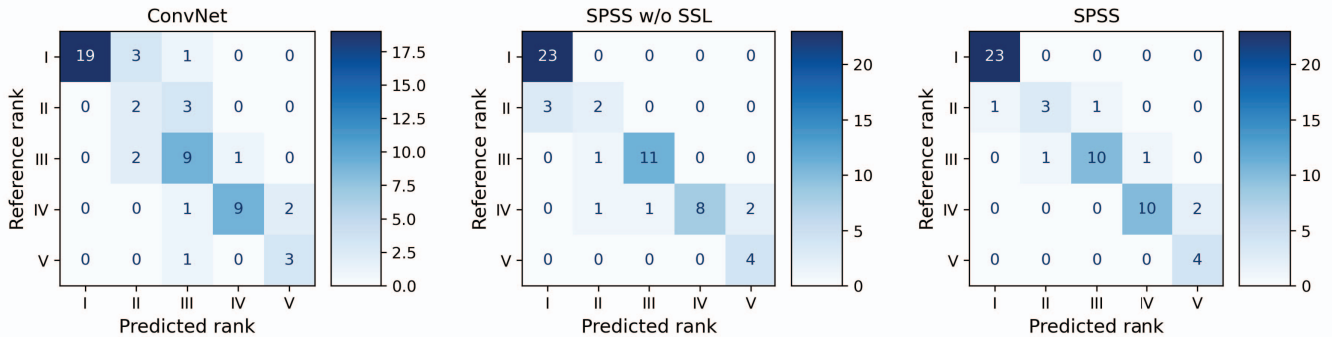

Figure 7. Confusion matrix of the three models on the non-gated CT test set. Ground truth scores are on the y-axis and model predictions are on the x-axis of each matrix.

the feature map. Therefore, $\mathcal{L}_{ssl}$ w/$\mathcal{L}_{ls}$ can optimize the mean absolute scoring bias to close to 50. These phenomena prove that $\mathcal{L}_{ls}$ can accelerate the convergence ability of the model and reduce scoring bias.

Scoring bias and consistency analysis. According to the CAC regions predicted by SPSS, we further calculate the CCS of per-artery by Agatston score [1]. The ME for per-arterial CCS are listed in Table 1 and Kappa for patient CVD ranks are listed in Figure 4.

Compared with SPSS w/o SSL, our SPSS achieves competitive ME and superior Kappa on both gated CT and non-gated CT. Regarding ME, our SPSS constructs the best ME (bold font) for LM, LCX, LAD and total calcium of gated CT. Biases were -2.62, 0.54, -0.11 and 2.31 for LM, LCX, LAD and total calcium, respectively. Qualitatively, CVD ranks (I–V, for CCS of 0, 1–10, 11–100, 101–400, >400, respectively) were evaluated using Cohen's Kappa statistic, our SPSS achieves the best agreement in gated CT (Kappa = 92.4%). It can be seen that the introduction of gated CT for training can enrich the information on CAC and improve the ability of the model to identify the CAC. For non-gated CT, our SPSS constructs the best ME for LM, LAD, RCA and total calcium compared to SPSS w/o SSL. Biases were -0.25, -12.30,-1.99 and -15.83 for LM, LAD, RCA and total CAC, respectively. In the CVD ranks, our SPSS achieved

the best Kappa of 85.2%. Additionally, we observed that the standard deviation of non-gated CT was generally higher compared to that of gated CT. The reason is that CAC regions are more difficult to identify on non-gated CT, which is due to the lower image quality and motion artifact interference of non-gated CT.

It is necessary to discuss such results from two aspects, including artifact countermeasures and CAC segmentation of per-arterial. As shown in Figure 5, since the non-gated CT dataset does not provide the real CAC mask, we only need to make a quantitative comparison with the reference CCS. For SPSS w/o SSL, since semantic supervised learning of gated CT does not participate in training, it is difficult for the model to identify CAC. Especially when there are many calcification artifacts in the image, SPSS w/o SSL tends to segment aortic calcification or miss-segment CAC, which will lead to significant errors in scoring bias. On the contrary, by introducing semantic supervised learning of gated CT to participate in training, SPSS can better identify the CAC of non-gated CT, thereby reducing the scoring bias caused by artifacts. Therefore, the total CCS of SPSS can be closer to the reference CCS. Further, Figure 6 shows that our SPSS has a better 95% confidence interval ([-122.34, 90.67]). Benefiting from the smaller scoring bias, our SPSS can more accurately classify CVD ranks II, III, and IV of

non-gated CT, as shown in Figure 7. Furthermore, Table 1 and Figure 5 demonstrate that our SPSS can provide more accurate CCS for per-artery. While this is not required for CVD ranks, it may be interesting for clinical research. At last, we can conclude that semantic supervised learning of gated CT can prompt regression supervised learning to learn CAC semantic information of non-gated CT, thereby reducing potential scoring bias.

Comparison with other methods. Table 1 shows a comparison with other state-of-the-art calcium scoring methods by ConvNet and SE-ResNeXt using the same datasets. The proposed method achieves state-of-the-art performance compared to these methods. ConvNet directly performs regression learning on the per-arterial CCS, which can determine the CCS faster. Nevertheless, performance was slightly poor when total calcium was determined. As shown in Figure 4 and Figure 6, ConvNet tends to overestimate CCS, making CVD rank I and II wrongly divided into III and IV. This difference in performance may be a consequence of the increased complexity of the per-artery scoring task. In addition, compared to SE-ResNeXt, since semantic supervised learning of gated CT does not participate in training, the Kappa of SPSS w/o SSL is low. Instead, by introducing semantic supervised learning of gated CT for training, the Kappa of our SPSS on non-gated CT is 1.6% more than SE-ResNeXt. This illustrates the semantic supervised learning of gated CT and the regression supervised learning of non-gated CT can better identify CAC and combat the plague of pixel-level artifact labels.

5. Conclusions

We have developed a novel Semantic-Prompt Scoring Siamese (SPSS) network for the coronary calcium scoring (CCS) of non-gated CT scans. In SPSS, we construct a shared network using regression supervised learning and semantic supervised learning. To train the SPSS network, we combine non-gated CT scans without calcium masks and gated CT scans with calcium masks. The SPSS network is trained using regression supervised learning to predict the CCS of the non-gated CT scans. To mitigate the impact of motion artifacts, gated CT scans are utilized to train the network to learn more accurate semantic features related to per-artery. The integration of regression supervised learning and semantic supervised learning enables the network to accurately predict the CCS of the non-gated CT scans based on the semantic information, thereby eliminating the need for pixel-wise artifact labels. By evaluating a public dataset, our SPSS establishes a state-of-the-art performance.

References

[1] Arthur S Agatston, Warren R Janowitz, Frank J Hildner, Noel R Zusmer, Manuel Viamonte Jr, and Robert Detrano. Quantification of coronary artery calcium using ultrafast computed tomography. *Journal of the American college of cardiology*, 15(4):827–832, 1990.

[2] Carlos Cano-Espinosa, Germán González, George R Washko, Miguel Cazorla, and Raúl San José Estépar. Automated agatston score computation in non-ecg gated ct scans using deep learning. In *Medical Imaging 2018: Image Processing*, volume 10574, pages 673–678. SPIE, 2018.

[3] Bob D de Vos, Jelmer M Wolterink, Tim Leiner, Pim A de Jong, Nikolas Lessmann, and Ivana Išgum. Direct automatic coronary calcium scoring in cardiac and chest ct. *IEEE transactions on medical imaging*, 38(9):2127–2138, 2019.

[4] Robert Detrano, Alan D Guerci, J Jeffrey Carr, Diane E Bild, Gregory Burke, Aaron R Folsom, Kiang Liu, Steven Shea, Moyses Szklo, David A Bluemke, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *New England Journal of Medicine*, 358(13):1336–1345, 2008.

[5] David Eng, Christopher Chute, Nishith Khandwala, Pranav Rajpurkar, Jin Long, Sam Shleifer, Mohamed H Khalaf, Alexander T Sandhu, Fatima Rodriguez, David J Maron, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. *NPJ digital medicine*, 4(1):88, 2021.

[6] Bernhard Föllmer, Federico Biavati, Christian Wald, Sebastian Stober, Jackie Ma, Marc Dewey, and Wojciech Samek. Active multitask learning with uncertainty-weighted loss for coronary calcium scoring. *Medical Physics*, 49(11):7262–7277, 2022.

[7] Harvey S Hecht. Coronary artery calcium scanning: past, present, and future. *JACC: Cardiovascular Imaging*, 8(5):579–596, 2015.

[8] Harvey S Hecht, Paul Cronin, Michael J Blaha, Matthew J Budoff, Ella A Kazerooni, Jagat Narula, David Yankelevitz, and Suhny Abbara. 2016 scct/str guidelines for coronary artery calcium scoring of noncontrast noncardiac chest ct scans: a report of the society of cardiovascular computed tomography and society of thoracic radiology. *Journal of cardiovascular computed tomography*, 11(1):74–84, 2017.

[9] Thijmen W Hokken, Joris F Ooms, Isabella Kardys, Alexander Hirsch, Annick C Weustink, Sanne Schipper, Peter Heil, Joost Daemen, Ricardo PJ Budde, and Nicolas M Van Mieghem. Validation of a three-dimensional computed tomography reconstruction tool for aortic valve calcium quantification. *Structural Heart*, 7(2):100122, 2023.

[10] Mohammad Arafat Hussain, Alborz Amir-Khalili, Ghassan Hamarneh, and Rafeef Abugharbieh. Segmentation-free kidney localization and volume estimation using aggregated orthogonal decision cnns. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pages 612–620. Springer, 2017.

[11] Peter C Jacobs, MJ Gondrie, Y Graaf, Harry J de Koning, Ivana Isgum, B van Ginneken, and WP Mali. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose ct screening for lung cancer. 2012.

[12] Peter C Jacobs, Mathias Prokop, Yolanda van der Graaf, Martijn J Gondrie, Kristel J Janssen, Harry J de Koning, Ivana Isgum, Rob J van Klaveren, Matthijs Oudkerk, Bram van Ginneken, et al. Comparing coronary artery calcium and thoracic aorta calcium for prediction of all-cause mortality and cardiovascular events on low-dose non-gated computed tomography in a high-risk population of heavy smokers. *Atherosclerosis*, 209(2):455–462, 2010.

[13] Zoé Lambert, Caroline Petitjean, Bernard Dubray, and Su Kuan. Segthor: Segmentation of thoracic organs at risk in ct images. In *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2020.

[14] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[15] Nikolas Lessmann, Bram van Ginneken, Majd Zreik, Pim A de Jong, Bob D de Vos, Max A Viergever, and Ivana Išgum. Automatic calcium scoring in low-dose chest ct using deep neural networks with dilated convolutions. *IEEE transactions on medical imaging*, 37(2):615–625, 2017.

[16] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[17] Haruhiko Machida, Isao Tanaka, Rika Fukui, Yun Shen, Takuya Ishikawa, Etsuko Tate, and Eiko Ueno. Current and novel imaging techniques in coronary ct. *Radiographics*, 35(4):991–1010, 2015.

[18] K Nasir, MS Bittencourt, MJ Blaha, R Blankstein, AS Agatson, JJ Rivera, Michael D Miedema, CT Sibley, LJ Shaw, RS Blumenthal, et al. Implications of coronary artery calcium testing among statin candidates according to american college of cardiology/american heart association cholesterol management guidelines: Mesa (multi-ethnic study of atherosclerosis)(vol 66, pg 1657, 2015). *Journal of the American College of Cardiology*, 2015.

[19] B Ohnesorge, T Flohr, R Fischbach, A Kopp, A Knez, S Schröder, U Schöpf, A Crispin, E Klotz, M Reiser, et al. Reproducibility of coronary calcium quantification in repeat examinations with retrospectively ecg-gated multisection spiral ct. *European Radiology*, 12:1532–1540, 2002.

[20] World Health Organization et al. Who package of essential noncommunicable (pen) disease interventions for primary health care. *WHO*, 2020.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[22] John A Rumberger, Bruce H Brundage, Daniel J Rader, and George Kondos. Electron beam computed tomographic coronary calcium scanning: a review and guidelines for use in asymptomatic persons. In *Mayo Clinic Proceedings*, pages 243–252. Elsevier, 1999.

[23] S Saluja, J Sobolewska, K Khan, H Contractor, LJ Mitchell, S Saluja, M Daniells, X Jiang, and SG Anderson. The clinical implications of incidental coronary artery calcification in routine, non-triggered high-resolution thoracic computed tomography: a retrospective study. *European Heart Journal*, 41(Supplement_2):ehaa946–0173, 2020.

[24] Gianmarco Santini, Daniele Della Latta, Nicola Martini, Gabriele Valvano, Andrea Gori, Andrea Ripoli, Carla L Susini, Luigi Landini, and Dante Chiappino. An automatic deep learning approach for coronary artery calcium segmentation. In *European Medical and Biological Engineering Confernce*, pages 374–377. Springer, 2017.

[25] Ran Shadmi, Victoria Mazo, Orna Bregman-Amitai, and Eldad Elnekave. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest ct. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 24–28. IEEE, 2018.

[26] Joseph Shemesh, Claudia I Henschke, Dorith Shaham, Rowena Yip, Ali O Farooqi, Matthew D Cham, Dorothy I McCauley, Mildred Chen, James P Smith, Daniel M Libby, et al. Ordinal scoring of coronary artery calcifications on low-dose ct scans of the chest is predictive of death from cardiovascular disease. *Radiology*, 257(2):541–548, 2010.

[27] Jane S Skinner, Liam Smeeth, Jason M Kendall, Philip C Adams, and Adam Timmis. Nice guidance. chest pain of recent onset: assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin. *Heart*, 96(12):974–978, 2010.

[28] Sanne GM van Velzen, Nikolas Lessmann, Birgitta K Velthuis, Ingrid EM Bank, Desiree HJG van den Bongard, Tim Leiner, Pim A de Jong, Wouter B Veldhuis, Adolfo Correa, James G Terry, et al. Deep learning for automatic calcium scoring in ct: validation using multiple cardiac ct and chest ct protocols. *Radiology*, 295(1):66–79, 2020.

[29] Wenjia Wang, Lin Yang, Sicong Wang, Qiong Wang, and Lei Xu. An automated quantification method for the agatston coronary artery calcium score on coronary computed tomography angiography. *Quantitative Imaging in Medicine and Surgery*, 12(3):1787, 2022.

[30] Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, and Shuo Li. Full left ventricle quantification via deep multitask relationships learning. *Medical image analysis*, 43:54–65, 2018.

[31] Joseph Yeboah, Robyn L McClelland, Tamar S Polonsky, Gregory L Burke, Christopher T Sibley, Daniel O'Leary, Jeffery J Carr, David C Goff, Philip Greenland, and David M Herrington. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. *Jama*, 308(8):788–795, 2012.

[32] Xiantong Zhen, Heye Zhang, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li. Direct and simultaneous estimation of cardiac four chamber volumes by multioutput sparse regression. *Medical image analysis*, 36:184–196, 2017.