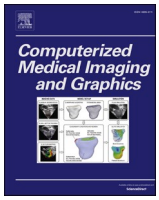
Computerized Medical Imaging and Graphics 121 (2025) 102503

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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



TQGDNet: Coronary artery calcium deposit detection on computed tomography

Wei-Chien Wang [a,1], Christopher Yu [b,c,d,1], Euijoon Ahn [e], Shahab Pathan [f], Kazuaki Negishi [g,h,i,j,k,2], Jinman Kim [a,l,*,2]

[a] Biomedical Data Analysis and Visualisation Lab, School of Computer Science, The University of Sydney, Australia
[b] Sydney Medical School, University of Sydney, Sydney, NSW, Australia
[c] University of Toronto, Toronto, Ontario, Canada
[d] Peter Munk Cardiac Centre, Toronto General Hospital, Toronto, Ontario, Canada
[e] College of Science and Engineering, James Cook University, Cairns, Australia
[f] Department of Cardiology, Nepean Hospital, Australia
[g] School of Clinical Medicine, University of New South Wales, Australia
[h] Department of Cardiology, Liverpool Hospital, Liverpool, NSW, Australia
[i] The Ingham Institute for Applied Medical Research, Sydney, NSW, Australia
[j] Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia
[k] Nepean Clinical School, University of Sydney, Australia
[l] Telehealth and Technology Centre, Nepean Hospital, Australia

ARTICLE INFO

Keywords:
Coronary artery calcium
Deep learning
Fusion modules
Detection

ABSTRACT

Coronary artery disease (CAD) continues to be a leading global cause of cardiovascular related mortality. The scoring of coronary artery calcium (CAC) using computer tomography (CT) images is a diagnostic instrument for evaluating the risk of asymptomatic individuals prone to atherosclerotic cardiovascular disease. State-of-the-art automated CAC scoring methods rely on large annotated datasets to train convolutional neural network (CNN) models. However, these methods do not integrate features across different levels and layers of the CNN, particularly in the lower layers where important information regarding small calcium regions are present. In this study, we propose a new CNN model specifically designed to effectively capture features associated with small regions and their surrounding areas in low-contrast CT images. Our model integrates a specifically designed low-contrast detection module and two fusion modules focusing on the lower layers of the network to connect more deeper and wider neurons (or nodes) across multiple adjacent levels. Our first module, called ThrConvs, includes three convolution blocks tailored to detecting objects in images characterized by low contrast. Following this, two fusion modules are introduced: (i) Queen-fusion (Qf), which introduces a cross-scale feature method to fuse features from multiple adjacent levels and layers and, (ii) lower-layer Gather-and-Distribute (GD) module, which focuses on learning comprehensive features associated with small-sized calcium deposits and their surroundings. We demonstrate superior performance of our model using the public OrCaScore dataset, encompassing 269 calcium deposits, surpassing the capabilities of previous state-of-the-art works. We demonstrate the enhanced performance of our approach, achieving a notable 2.3–3.6 % improvement in mean Pixel Accuracy (mPA) on both the private Concord dataset and the public OrCaScore dataset, surpassing the capabilities of established detection methods.

1. Introduction

Cardiovascular disease (CVD) stands as one of the main global causes of mortality, claiming around 17.9 million lives yearly, as reported by the World Health Organization (WHO) (World Health Organization, 2021). Among the various CVDs, atherosclerotic CVD (ASCVD) is the

* Corresponding author at: Biomedical Data Analysis and Visualisation Lab, School of Computer Science, The University of Sydney, Australia.
E-mail address: jinman.kim@sydney.edu.au (J. Kim).
[1] co-first author
[2] co-last author

https://doi.org/10.1016/j.compmedimag.2025.102503
Received 25 August 2024; Received in revised form 23 December 2024; Accepted 25 January 2025
Available online 6 February 2025

most widespread and deadly form. This condition is associated with the accumulation of calcium deposits in the coronary arteries. Recent guidelines and research by the American Heart Association (AHA) indicate that employing coronary artery calcium (CAC) scoring can proficiently stratify patient groups who are at risk and assist in making appropriate treatment decisions, particularly for those falling into the Intermediate Risk category with a 5–10 % risk of ASCVD over a 10-year period (Naghavi et al., 2019; Gulati et al., 2021). As part of the CAC scoring process, quantification of calcium deposit accumulations observable on Computer Tomography (CT) images, is required. The calcium deposits are generally quantified via detection and/or segmentation methods. Within these images, electrocardiogram (ECG)--gating involves acquiring images synchronously with ECG signal to correct for the effects of cardiac motion by timing image acquisition to specific phases of the cardiac cycle. Non-gated CT are also routinely acquired, but comparative to ECG-gating, there are more noise and artefacts from cardiac motion, resulting in blurry and lower-quality images. Non-gated CT scans are more easily obtained than gated ones for several reasons. First, they are quicker to conduct, making them more appropriate for emergency situations or patients who may struggle with holding their breath which is necessary for gated CT. Additionally, they are more cost-effective due to requiring less expertise and specialized equipment. Moreover, non-gated scans alleviate patient discomfort by not requiring synchronization with the heartbeat, thus decreasing the probability of motion artifacts.

The quantification of calcium deposits for CAC scoring is generally done by clinicians which is an expensive, complex, labour-intensive process, and is also susceptible to human subjectivity and errors. Hence, automatic quantification of CAC is an important task to aid in treatment decisions. The calcium deposits, typically comprising of few pixels per image slice as depicted in Fig. 1, present three unique challenges: 1) Calcium deposits often have similar intensity values to its surrounding structures, such as with pericardium or epicardial fat; 2) Calcium deposits, especially those acquired without ECG gating, are susceptible to interference from noise, motion artifacts, or other disturbances in the image, such as metallic artifacts, beam hardening and scatter radiation and, 3) Calcium deposits may be too small to have sufficient spatial resolution to capture relevant information, especially in the early stages of calcium formation. These challenges may lead to a lack of detail or ambiguity in quantifying the deposits.

There has been sustained efforts in trying to automatically quantify CAC scores. Early studies used multi-atlas-based methods (Isgum et al., 2009, 2012; Takx et al., 2014). For example, an "atlas" or pre-labelled reference image, was used to serve as a model to automatically segment regions of interest (ROIs) in the images. Then, various metrics such as volume, intensity, or density can be calculated within the ROIs to quantify the calcium deposits. In addition to multi-atlas registration, Ding et al. (2015) also incorporated knowledge-based vessel region separation algorithms for both global and local CAC scoring. Following the segmentation of essential heart structures, including the heart region, ascending aorta, left and right ventricles, and aortic root, the algorithm further delineates the coronary artery territories, such as the right coronary artery (RCA), left circumflex artery (LCA), and left anterior descending artery (LAD). It then applies region growing techniques to identify calcifications within these targeted areas.

However, a major drawback with the use of multiple atlases is the reliance on accurately registering the CT images to the atlas. Alternatively, Isgum et al. (2007) characterized the position of calcification lesions in relation to the heart and the aorta using a set of approximated Gaussian derivatives, followed by three classification strategies, including two sequential k-nearest neighbor (kNN) classifiers, to identify the lesions. Kurkure et al. (2010) implemented a hierarchical approach to firstly detect the peripheral structure of the heart, then determining it through a combination of prior anatomical knowledge and dynamic programming. Within the acquired heart regions, they introduced a hierarchical classifier to differentiate arterial calcifications (both coronary and aortic) from other candidate regions, and then to further distinguish between coronary and aortic calcifications.

Recently, with the emergence of artificial intelligence (AI), in particular, deep learning (DL) techniques, they have been broadly applied to assist in automatic CAC scoring (Li et al., 2020; You et al., 2018; Ronneberger et al., 2015; Kang et al., 2019). Lessmann et al (Lessmann et al., 2018). employed two consecutive Convolutional Neural Networks (CNNs) to detect coronary artery / cardiac valve calcifications and thoracic aorta. They initially utilized a dilated CNN with a large receptive field to anatomically locate potential calcifications. The subsequent CNN (Lessmann et al., 2016) was then developed to enhance the results of the first CNN by differentiating true calcifications from false positives using local information that shares similar appearance and location characteristics. van Velzen et al. (2020) further extended

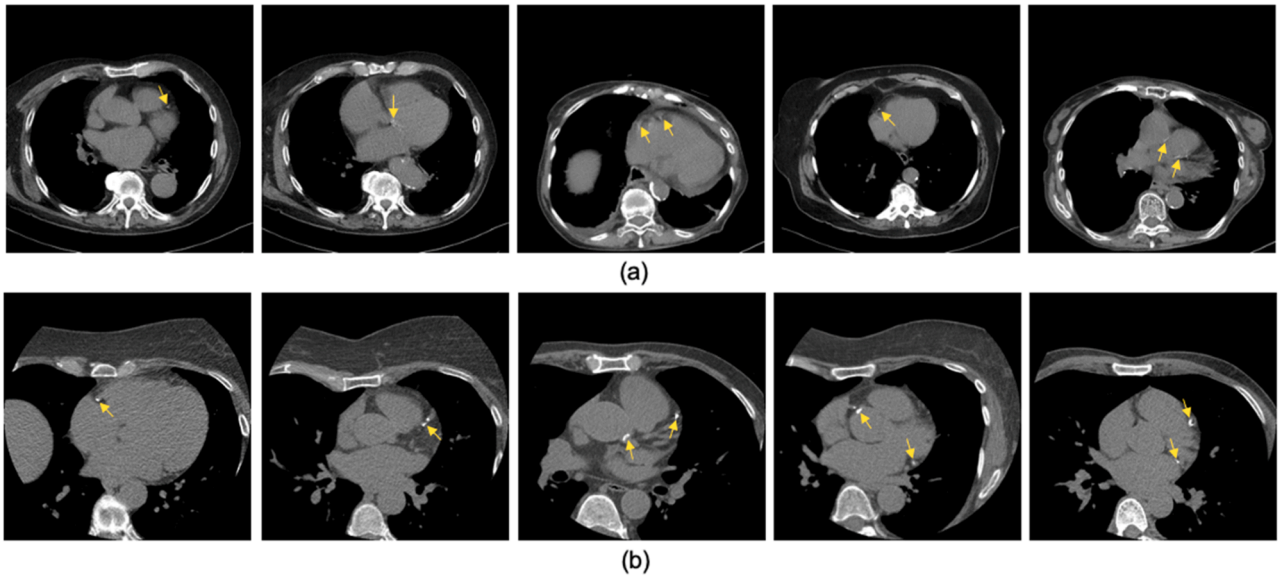


Fig. 1. Example images of calcium deposit in both gated and non-gated CT images. The top row (a) is from our Concord dataset (non-gated) and the bottom row (b) is the public OrCaScore dataset (gated). The calcium regions are indicated by yellow arrows. Different imaging resolutions can be seen between our Concord dataset (a) and the public dataset (b). The public dataset had a more focused field of view (focused on the heart), and better contrast for calcium differentiation.

the research of Lessmann et al. (2018) by evaluating the efficacy of automated coronary artery calcium (CAC) scoring across a variety of cardiac-related CT examinations. These included CAC screening CT, diagnostic chest CT, radiation therapy planning CT, coronary artery calcium (CAC) scoring CT, PET attenuation correction CT, and low-dose chest CT. In this context, 'diverse' refers to the different types of CT scans used to assess the AI model's adaptability and performance across various medical imaging scenarios. van den Oever et al. (2020) conducted research on the effect of dilated convolutional layers within the CNN architectures on CAC and concluded that these layers resulted in improvements in sensitivity, specificity, precision, and negative predictive values. In addition, some research has focused on segmenting calcium deposits using specific CNN networks, such as ResNet (Chamberlin et al., 2021; Eng et al., 2021; Peng et al., 2023) and VGG (Klug et al., 2022). Datong et al. (2019) employed ResNet-50 as the benchmark network for the Single Shot MultiBox Detector (SSD) with an aggregate channel feature model to identify suspected calcium areas, thereby significantly reducing the time required for single-frame image detection. Moreover, some studies initially cropped the heart region using the cardiac atlas (Santini et al., 2017) or with an anatomical model (Rasul et al., 2023), and subsequently segmented the calcium deposits within the cropped region. In another study, Lee et al. (2021) initially cropped the heart section from the CT images and divided it into nine sub-images, enlarging them to serve as input for comparing the performance of three CNNs: Inception (Szegedy et al., 2015), ResNet and VGG and found that the ResNet50 model produced the best results. In the study by Zhang et al. (2018), they employed joint learning of 2D (DenseNet and U-Net) modules to extract intra-slice calcification features and 3D (3D U-Net) modules to extract inter-slice calcification features. This approach extracted rich semantic features for artery-specific calcification identification, establishing it as a reliable clinical diagnostic method for detecting coronary calcifications based on quantitative results.

Despite the advances in calcium segmentation / detection for CAC scoring, existing methods continue to struggle with the difficulty from small calcium regions. This challenge is not exclusive to CAC scoring and there have been several studies aimed at better detecting small regions, such as with lesions among various medical image modalities. As an example, CNNs were enhanced with the integration of channel-wise parallel attention block and bi-directional spatial attention block (Bhati et al., 2024) to extract features for the detection of small diabetic retinopathy lesions in retinal fundus images. In another study, Ahmad et al. (2023) enhanced the identification of small gastric lesions in endoscopic images by incorporating a Squeeze-and-Excitation attention block into YOLOv7. Compared to YOLOv5, YOLOv7 adopts a deeper network architecture with additional convolutional layers and residual blocks, while reducing the number of fully connected layers and hence decreasing the parameters by 40 % and computational load by 50 %. Jiang et al. (2023) integrated a multi-head self-attention module and a Shuffle Attention module into YOLOv5s to enhance the identification of lymphocytic infiltrative lesions in pathological images.

These studies all relied on the use of an attention block which has been demonstrated for its ability to aid in small region detection. Another common technique is to fuse features from different CNN levels and layers to better capture characteristics of small lesions (Jia et al., 2018; Le et al., 2020; Elhanashi et al., 2023). In another approach, Li et al. (2021) improved the contrast between normal brain tissue and hemorrhagic areas by combining the original CT slice with its flipped counterpart, thereby introducing symmetry constraints for brain images in the proposed model. These small region techniques, to our knowledge, has not been exploited and optimized for CAC scoring.

In this study, we propose a novel method termed TQGDNet (ThrConvs-Queen fusion-Gather-and-Distribute fusion) for automated detection of CAC in CT images. Detection is a critical initial step toward CAC scoring. Our method integrates three distinct modules, as shown in Fig. 3: 1) three-convolutional module, ThrConvs, designed to identify

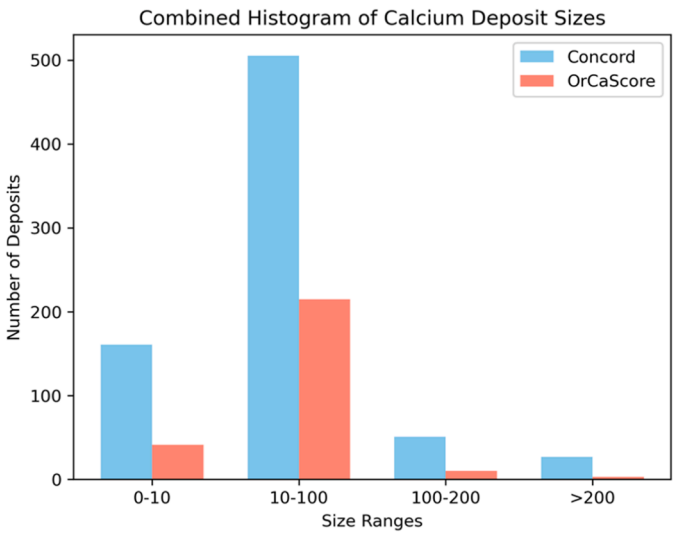


Fig. 2. The figure illustrates the distribution of calcium deposit sizes across the Concord and OrCaScore datasets. The x-axis (Size Ranges) categorizes the calcium deposits into four groups: 0–10 pixels, 10–100 pixels, 100–200 pixels, and $>$ 200 pixels. The y-axis (Number of Deposits) indicates the number of calcium deposits (counts).

objects in low-contrast CT images. Inspired by the back-projection concept (Irani and Peleg, 1991), as implemented by Haris et al. (2018) in their work on learning across different resolutions, low-resolution data and residual connections were adopted to design the ThrConvs modules, specifically tailored for detecting small targets in medical images. Back-projection (Irani and Peleg, 1991) is a widely recognized iterative process for reducing reconstruction errors efficiently. Its effectiveness has been demonstrated in numerous studies (Zhao et al., 2017; Haris et al., 2017; Dong et al., 2009; Timofte et al., 2016). While originally developed for scenarios with multiple low-resolution (LR) inputs, back-projection can also be applied with a single LR image by iteratively upsampling the input using various upsampling operators and calculating the reconstruction error (Dai et al., 2007). Timofte et al. (2016) highlighted its ability to enhance the quality of super-resolution (SR) images, and Zhao et al. (2017) introduced a method leveraging iterative projection to refine high-frequency texture details. To enhance this algorithm, the ThrConvs module is designed to guide low-resolution (LR) tasks by employing interconnected two- and three-layer convolutional operations integrated with residual connections. Specifically, the relationship between LR images is constructed through iterative residual connections linking the second and third convolutional layers. This design allows the network to effectively retain essential LR components by utilizing learned convolutional and residual operators. As a result, the module generates deeper feature representations, supporting the construction of diverse and detailed LR features. ThrConvs compares derived features from the three convolutional viewpoints—specifically, the features obtained after the first, second, and third convolutional calculations—to capture detailed information about the calcium region and their surroundings; 2) A cross-scale fusion module, termed the Queen fusion (Qf), integrates features from adjacent layers and levels of the ThrConvs and, 3) Gather-and-Distribute (GD) fusion module which captures the features collected from the lower layers of the network to improve the characterization of small calcium regions. We evaluated our method using both Concord and public calcium datasets, encompassing gated and non-gated data. This diversity allows us to demonstrate how our method can effectively handle different types of cardiac imaging modalities. Subsequently, we benchmarked its performance against state-of-the-art calcium detection and segmentation methods.

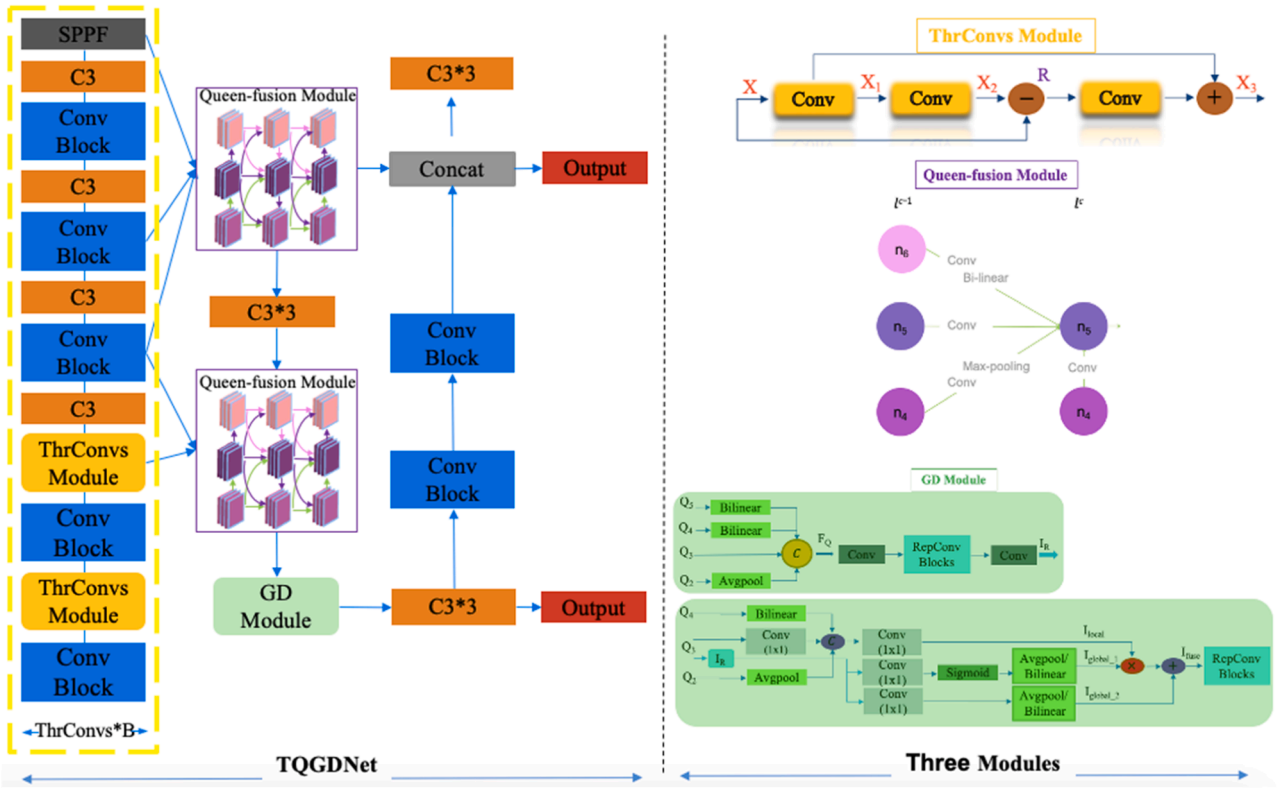


Fig. 3. The proposed TQGDNet method comprising of three modules (left of the figure): ThrConvs (yellow), Queen fusion (purple) and GD fusion (green). The detailed architectures of each module are presented (on the right side). ThrConvs*B, shown inside the yellow dotted line, is the backbone of the TQGDNet that integrates our ThrConvs.

2. Methods

2.1. Materials

2.1.1. Concord Concord Dataset (Non-Gated)

Our Concord dataset (Yu et al., 2021), obtained from the Department of Cardiology at Concord Repatriation General Hospital, includes 32 patients (51 % male, mean age: 69 years). The dataset was acquired using non-gated, non-contrast-enhanced CT chest imaging. Further details regarding the dataset are available in (Yu et al., 2021). The requirement for informed consent was waived by the local Ethics Committee. CT scans were acquired using a Definition DS 64-slice or Definition AS+ 128-slice scanner (Siemens, Germany). All CT scans were performed for non-cardiac reasons. Calcified lesions were identified based on a minimum threshold of 130 Hounsfield units. CT scans were evaluated by a single observer who was blinded to the clinical outcomes. The CT images were labelled by a clinician to indicate the calcium regions by placing bounding boxes around them using the LabelImg (Tzutalin, 2015) software. For each patient, there were 1–20 image slices with identified calcium deposits, totalling 257 images. The dataset was split into 4:1 ratio, allocating 205–206 images for training (from 26 patients) and reserving 51–53 images for testing (from 6 patients).

2.1.2. Public OrcaScore Dataset (Gated)

The OrCaScore challenge dataset (Purpose, 2016) was obtained from four academic hospitals: Antwerp University Hospital (Antwerp, Belgium), Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands), University Medical Center Groningen (Groningen, The Netherlands), and University Medical Center Utrecht (Utrecht, The Netherlands). Imaging was conducted using four different CT scanners, with acquisitions synchronized to the diastolic rest period through ECG-triggering at 70 % (GE, Siemens), 75 % (Toshiba), or 78 % (Philips). The dataset includes 32 patients (50 % male). Further details regarding the dataset can be found in Purpose (2016). A 130 Hounsfield Units (HU) threshold was applied to identify potential coronary artery calcium (CAC) lesions. Annotations were performed using custom software (iX Viewer, Utrecht, The Netherlands) by a research physician with 5 years of experience and a radiologist with 12 years of expertise. Each training exam comprised between 1 and 20 calcium-containing images, resulting in a total of 163 images. The dataset was split into training and testing sets using a 4:1 ratio, with 131–132 calcium-containing images from 24 patients assigned to the training set, and 32–33 images from 8 patients allocated to the testing set. To ensure consistency in ground truth labels between the OrCaScore dataset and the Concord dataset, pixel-level annotations from the OrCaScore dataset were transformed into detection-level annotations by assigning bounding boxes to the lesions.

2.1.3. Key differences between the datasets

In our Concord dataset, due to non-gated imaging, there was narrower contrast between calcium deposits and the background compared to the public dataset. Detecting calcium on a less pronounced contrast background makes it more challenging. Fig. 1 illustrates some examples from both the public dataset and our Concord dataset. In regard to the distribution of calcium deposit sizes, the two datasets vary greatly with sizes for Concord and public data as depicted in Fig. 2. The Concord dataset poses a greater challenge in calcium detection due to its lower contrast and resolution with smaller calcium deposits sizes compared to the public dataset. Furthermore, the inclusion of both gated and non-gated images diversifies the training data, making the model more robust and adaptable to various clinical scenarios. We suggest that the variability in patient demographics and imaging conditions mitigates the risk of overfitting to a specific dataset type, thereby enhancing the

model's ability to generalize to unseen data. In overall, As shown in Fig. 2, majority of calcium deposits in the public dataset fall within the 10–100 size range, providing sufficient feature information for effective detection. However, the Concord dataset poses greater challenges, with a higher proportion of 0–10 size calcium deposits. These deposits are particularly underrepresented and difficult to identify due to the low-contrast imaging conditions.

2.1.4. Data pre-processing

For the pre-processing of OrCaScore scans, the raw data was read from files in mhd and zraw formats. The Hounsfield Unit (HU) values were clipped to a range of -1200 -1200 to 18001800 to focus on relevant anatomical structures while minimizing the influence of outliers. Subsequently, an automated cropping step was applied to isolate the heart region using a pre-trained model. This model was trained on a dataset of 300 images annotated with bounding boxes specifically marking the heart. For bounding box labels, to meet the diverse input requirements of state-of-the-art methods, the masks were converted into YOLO format, which specifies normalized bounding box coordinates, and XML format (vol format) for compatibility with alternative frameworks. For the Concord scans, images were loaded from a dataset in PNG format. Similar to the OrCaScore scans, an automated heart cropping step was performed using the same pre-trained model, ensuring consistent localization of the heart across different data modalities. Additionally, the masks were transformed into YOLO format and XML format to accommodate various model architectures and analysis workflows. This standardized preprocessing pipeline ensures compatibility and optimizes data preparation for subsequent analytical tasks.

2.2. TQGDNet

2.2.1. The ThrConvs module

The "ThrConvs" module enhances feature learning by performing addition and subtraction operations after one to three convolutional neural networks. Inspired by the work of Haris et al. (2018), who utilized deconvolution and convolution operations to learn features across resolution. However, this process could sacrifice detailed information about small targets within the receptive field, which expands during the transition from low-resolution to high-resolution. Therefore, we replaced the deconvolution layers used by Haris et al. (2018) with convolution layers, as depicted in Fig. 4. From the perspective of the three convolutions, performing subtraction computations between the original input and the output after two convolutions could enhance the learning of subtle information in objects. Subsequently, applying one additional convolution operation and then combining its output with that of the previous one-operation convolution helps retain key information. The formula can be defined as follows:

$$X_1 = Conv(X) \tag{1}$$

$$X_2 = Conv(X_1) \tag{2}$$

$$R = X - X_2 \tag{3}$$

$$X_3 = Conv(R) + X_1 \tag{4}$$

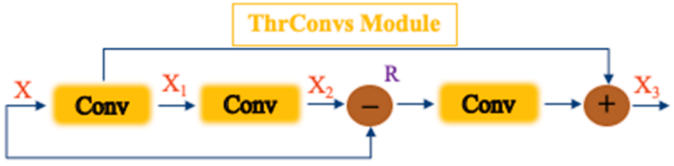


Fig. 4. The ThrConvs module employs addition and subtraction within three convolutions to concentrate on learning features of subtle calcium from low-contrast background images.

where X is the input, X_1represents the outcome after the first convolution operation, while X_2represents the result after two consecutive convolution operations with filter sizes of 8×8 and 12×12. R represents the residual between X and X_2. X_3 is the sum of the convolution operation of R and X_1. The subtraction in the ThrConvs module enables the removal of unusual features while preserving the common features present in most slices. We integrated the ThrConvs module at the beginning of the YOLO backbone (Jocher, 2020), as depicted inside the yellow dashed box in Fig. 3 and named it as "ThrConvs*B" with "B" shortening of backbone. In other words, ThrConvs*B serves as the backbone of TQGDNet.The C3 module employs three convolutional layers integrated with residual connections to improve the feature extraction. The initial two convolutional layers utilize asymmetric kernels to reduce channel dimensions and compress information, and in the process, enabling the network to capture a broader global context while minimizing computational demands. Subsequently, the final convolutional layer restores the channel dimensions to their original size, producing an output that retains the same spatial dimensions but with potentially altered channel characteristics. The term $C3 \times 3$ refers to the application of the C3 module three consecutive times.

2.2.2. Fusion modules

We implemented two kinds of fusion algorithms for two different purposes. Qf is designed to collect more extensive features across different levels and layers, while GD is designed for concentrating on searching for small calcium deposits. Firstly, to facilitate learning across different levels and layers, certain fusion features were introduced. The state-of-the-art fusion methods, as illustrated in Fig. 5, Feature Pyramid Network (FPN) (Lin et al., 2017) implements a one-way, top-down information flow within the same layers. However, to extract information from distant layers, it must undergo a 'recursive' procedure involving the collection of features from the previous two layers first and then the preceding layer. The transmission of features through this 'recursive' process in traditional FPN structures results in information loss. Due to the inherent limitation of this one-way flow of information, PANet (Jia et al., 2018), as employed by YOLOv5, incorporates an additional bottom-up path aggregation network into the FPN. Previous studies on combining features between neighboring layers only considered features at the same level (Jia et al., 2018) or the previous level (Le et al., 2020), which did not adequately account for capturing multi-scale features. However, to comprehensively integrate features from multiple levels and layers more effectively, we introduced a new cross-scale fusion mechanism, Qf. Secondly, CNNs have been successful in detecting smaller objects using lower-level features within the networks (Vapnik, 1998). CNNs operate on the principle that neurons in deeper layers boast larger receptive fields, facilitating the capture of extensive contextual information. This intrinsic trait makes these features highly proficient in detecting larger objects or overall structures within an image. On the contrary, neurons in the shallower layers of CNNs feature smaller receptive fields, allowing them to capture finer details in localized areas of the image. This characteristic enhances the suitability of these features for detecting smaller objects or local structures. In summary, higher-level features are employed to detect larger calcium deposits, whereas lower-level features are utilized for detecting smaller calcium deposits.

2.2.3. Queen-fusion (Qf)

The calcium dataset exhibited large-scale variance issues, with the ratio between the smallest objects and images (10/40000) and the largest objects and images (200/40000) being 1/20. Alleviating the issues caused by large-scale variations was challenging. Jiang et al. (2022) proposed a new cross-scale fusion technique called "Queen-fusion" to concatenate multi-scale features from more neighboring nodes in the previous and current layers. Queen-fusion, which takes into account features from both the same level and neighboring levels, similar to how

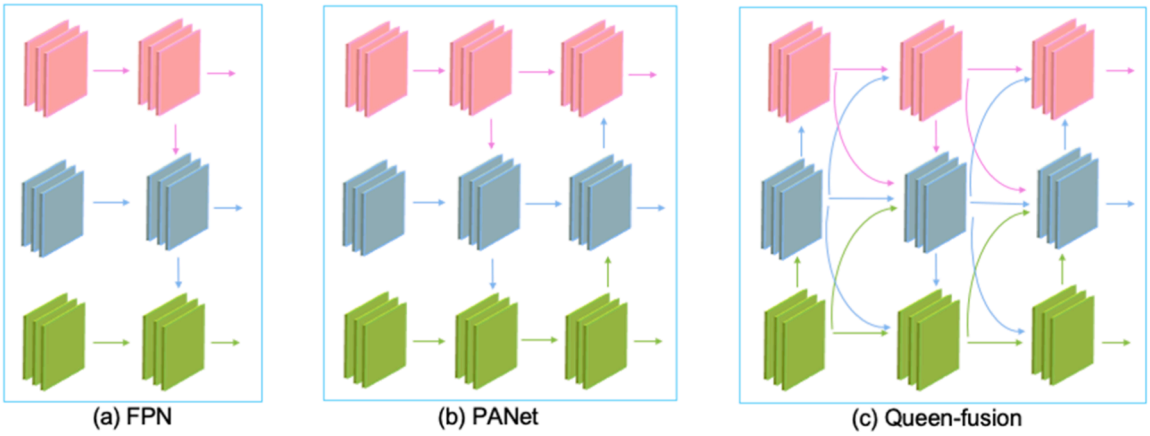


Fig. 5. Illustration of two state-of-the-art fusion methods and QF. (a) FPN (Lin et al., 2017) presents a top-down pathway for fusing multi-scale features; (b) PANet (Jia et al., 2018) incorporates an additional bottom-up pathway onto FPN and, (c) Queen-fusion aggregates features from more adjacent levels and layers.

a queen piece moves in chess. This encouraged the exchange of feature information among different spatial scales in adjacent layers. Queen-fusion was originally designed to tackle the large-scale variation issues present in the COCO public dataset. According to Singh and Davis (2018) the scale of the smallest and largest 10 % of object instances in the COCO public dataset is approximately 0.024 and 0.472, respectively. This represents an almost 20-fold difference in scale. In addition, in the COCO dataset, many object instances occupy less than 1 % of the image area, making it challenging to detect them.

To address the large-scale variation, it is necessary to consider more comprehensive information from various adjacent levels and layers. Therefore, we substituted the initial two layers with Queen-fusion in the YOLO backbone to tackle the challenges arising from the significant variation in object sizes relative to the images. This can be defined as follows:

$$n_5^{lc} = n_6^{lc-1} + n_5^{lc-1} + n_4^{lc-1} + n_4^{lc} \tag{5}$$

As shown in Fig. 6, node n_5 in layer *lc* concatenates four inputs: three from the previous layer, $lc-1$, and one from the current layer, *lc*. In $lc-1$, it includes up-sampling of node n_6, note n_5, and down-sampling of node n_4. At the same level, *lc*, there was only node n_4. The up-sampling used bilinear interpolation, while the down-sampling utilized max-pooling.

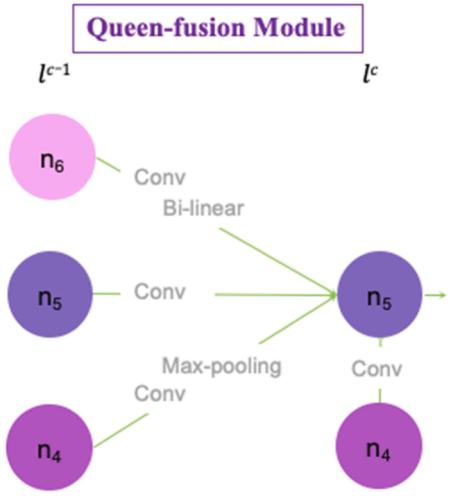


Fig. 6. Queen-fusion was employed to collect more feature information from distinct layers, encompassing lc and lc-1, as well as from various levels, such as n_4, n_5, and n_6.

2.2.4. Portion of gather-and-distribute module (GD)

To improve the model's capability in detecting objects of different sizes, Wang et al. (2024) developed two branches of GD: a low-stage GD primarily focusing on small and medium-sized objects, and a high-stage GD primarily targeted detecting large-sized objects. On both calcium CT datasets, our targets, calcium, occupy only a 1/1000 ratio of input images and object area, indicating that they belong to the category of small objects. Therefore, we placed greater emphasis on detecting small targets by implementing a lower-stage GD. More specifically, feature representations were gathered (G) and aligned from multiple levels by a unified module, and then these multiple aligned features were combined to produce a global feature, as shown in Fig. 7(a). This is followed by distributing (D) the acquired global feature across each level, as depicted in Fig. 7(b). It is noteworthy that, for the purpose of enhancing information flow, the features are combined from neighboring levels on a local scale. This approach not only prevents the loss of information found in the FPN structure but also improves the neck's ability to partially fuse information without introducing significant latency. The GD formula is as follows:

$$F_Q = AvgPool/Bilinear([Q_2, Q_3, Q_4, Q_5]) \tag{6}$$

$$I_R = RepConvBlocks(F_Q) \tag{7}$$

$$I_{global\_1} = resize(Sigmoid(Conv(I_R))) \tag{8}$$

$$I_{global\_2} = resize(Conv(I_R)) \tag{9}$$

$$I_{fuse} = Conv(Q_i) \times I_{global\_1} + I_{global\_2} \tag{10}$$

where the output feature maps, Q_2, Q_3, Q_4 and Q_5 from the Qf, were used for fusion to capture high-resolution features that preserve information about small targets. F_Q was derived by resizing the feature maps using average pooling and bilinear interpolation technologies to match the smallest feature size among those four feature maps. When using F_Q as the input for the multi-layer reparametrized convolutional blocks (RepConvBlock) to produce I_R, two different Convs are then employed with I_R to generate two sets of global information, represented as I_{global_1} and I_{global_2}. Qi represents the local information from the current level, and $Conv(Q_i)$was denoted as l_{local} in Fig. 7(b). At the end of the injection model, a RepConvBlock was incorporated to enable additional information extraction and fusion.

2.3. Three state-of-the-art methods

We compared our TQGDNet with three state-of-the-art optimized methods for calcium detection/segmentation methods. To maintain

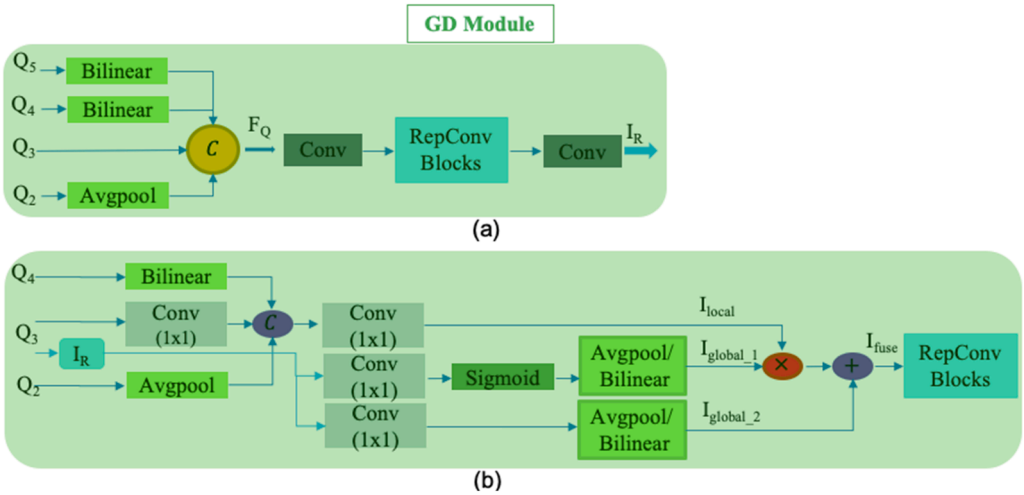


Fig. 7. The lower-layer Gather-and-Distribute (GD) branch comprising of two stages. The top row (a) depicts the gathering process, where feature maps from different levels (Q2, Q3, Q4, Q5) are assembled and aligned to a unified size. Subsequent to this, those aligned feature maps are fused using reparametrized convolutional blocks and convolutional blocks to produce global information. The bottom row (b) shows the aggregated global information received from the gathering process being distributed across each level to enhance detection capability.

consistency in the ground truth labels between our Concord dataset and the OrCaScore dataset, we converted the pixel-level labels in the OrCaScore dataset into detection-level labels by enclosing the lesions within bounding boxes. All competing methods used the ECG-gated dataset and the OrCaScore dataset as part of their experiment. Firstly, Zhao et al. (2020) used Faster R-CNN to detect coronary artery calcified plaque solely on the OrCaScore dataset, integrating medical prior knowledge through a proposed data augmentation technique into the training dataset. In addition, to decrease the creation of redundant anchor boxes, the region proposal networks (RPN) used in the original Faster R-CNN were replaced by guided anchoring (Wang et al., 2019). Secondly, Follmer et al. (2022) proposed an uncertainty-weighted multitask learning (MTL) model for coronary calcium scoring on three datasets, including DISCHARGE (Follmer et al., 2022), CADMAN (Dewey et al., n.d), and OrCaScore. More specifically, they jointly learned the multiclass coronary artery regions segmentation task and the binary lesion segmentation task to share complementary information. They also employed a multi-loss uncertainty-weighted (Kendall et al., 2017) approach to optimize the model parameters jointly. Thirdly, Zair et al. (2023) compared three different CNN-based networks, including U-Net, VGG16, and SegNet-VGG16 with transfer learning, for the segmentation of coronary artery calcification. They also compared images from the OrCaScore dataset with the heart area removed and retained.

3. Experiments and results

3.1. Implementation details

As a preprocessing step, we automatically cropped the images to the heart section using the YOLOv5 (Jocher, 2020) detection model. The detection model was trained using 300 images of our Concord dataset, which are distinct from the 257 images for calcium detection. Each of the images were accompanied by heart-bounding boxes labelled by a clinician. The trained model was then applied to both the Concord and public datasets.

The TQGDNet was trained from scratch for 200 epochs with a batch size of 4. We utilized the Stochastic Gradient Descent (SGD) optimizer to minimize the overall loss. We combined three loss functions (Redmon et al., 2016), including Objectness Loss, Localization Loss and, Classification Loss. The Objectness Loss, employing BCE (Binary Cross-Entropy) loss, assesses the disparity between the expected probability of calcium being enclosed within a bounding box and the actual probability of its presence therein. The Classification Loss, also using BCE loss, evaluates the discrepancy between the predicted class probability and the true class probability regarding its containment within the bounding box. The Localization Loss, employing GIOU (Generalized Intersection over Union Loss), assesses the disparity in coordinates between the predicted bounding box and the ground truth bounding box. The training commenced with a learning rate of 0.01 and concluded with a learning rate of 0.0001, coupled with a weight decay constant of 0.0005. We initiated a warm-up momentum that began at 0.8 for the first three epochs, and subsequently maintained the momentum at 0.937. We applied image data augmentation, incorporating operations like left-right flips, HSV augmentation (Hue: 0.015, Saturation: 0, Value: 0.4), image translation with a fraction of 0.1, and image mosaics. Both training and testing were conducted using a 12 GB NVIDIA GeForce RTX 2080 Ti. More specifically, dynamic learning strategies starting with a learning rate of 0.01 facilitate efficient exploration of the parameter space, enabling the model to identify potentially optimal regions during the initial training stages. Gradually reducing the learning rate to 0.0001 allows for fine-tuning and stabilization in later stages, preventing overshooting and enhancing convergence. This approach is particularly critical for small object detection, which requires precise localization and accurate feature representation. Additionally, the progressive decrease in the learning rate supports the refinement of the model's ability to extract subtle and intricate characteristics of small objects. This refinement enhances the delineation of fine-grained features, ultimately improving the accuracy of bounding box predictions and detection outcomes. Moreover, data Augmentation: To further enhance model performance, we employed data augmentation techniques commonly applied to medical images. These included left-right flips, HSV augmentation, image translation by a fractional amount, and image mosaics. These augmentations not only improve the model's generalization capabilities but also simulate variability in the data, contributing to better detection robustness.

3.2. Evaluation setup

Our evaluation employed multiple performance metrics, including precision, recall, mean Average Precision (mAP), and F1-score (F1). To ensure robust and reliable results, a 5-fold cross-validation was conducted on both the public OrCaScore dataset and the private Concord dataset (for additional experimental details, see Affiliate).

3.3. Comparison to the State-of-the-Art

Our proposed TQGDNet demonstrated superior performance compared to existing methods, achieving significant improvements on both the Concord and OrCaScore datasets. On the Concord dataset, TQGDNet outperformed the second-best method, Zhao et al. (2020) using Faster R-CNN, with a notable increase of 3.6 % in mAP. Similarly, on the public OrCaScore dataset, TQGDNet surpassed the second-best method, Zair et al. (2023) using SegNet-VGG16, achieving an improvement of 2.3 % in mAP. It is important to note that detection-based models such as YOLO and Faster R-CNN consistently exhibited higher precision than recall on both datasets. This is evident in the performance of Zhao et al. (2020) (Faster R-CNN-based) and TQGDNet (YOLO-based), which achieved the highest precision on the Concord and OrCaScore datasets, respectively. In contrast, Follmer et al. (2022) using MTL achieved the highest recall across both datasets. It is worth noting that the results on the public OrCaScore dataset were superior to those on the Concord dataset. This discrepancy can be attributed to differences in image complexity, including the higher prevalence of smaller calcium deposits (as illustrated in Fig. 2) and variations in image quality. Overall, TQGDNet exhibited strong performance on the F1-score, which considers both recall and precision, as shown in Table 1.

3.4. Ablation results

We conducted an ablation study to examine the individual contributions of each of the modules in the TQGDNet. Table 2 shows that the prediction performance was enhanced with the addition of two additional fusion modules to two backbones: one is YOLOv5*B, the backbone of YOLOv5, and the other one is ThrConvs*B, the backbone after adding ThrConvs modules to YOLOv5*B. In general, the addition of Qf and GD modules separately to ThrConvs*B resulted in greater improvements in prediction results on the Concord dataset compared to the public dataset. In particular, after adding the GD module, assistance to ThrConvs*B increased the mAP by 4.1 % in the Concord dataset and 3.4 % in the public dataset.

We further investigated the efficiency of adding these two fusion modules by incorporating them into YOLOv5*B, as depicted in Table 2. We observed that adding the GD module to YOLOv5B significantly improved the mAP by more than 3.5 % on both datasets. However, the contribution of adding GD to ThrConvsB was slightly greater compared to adding GD to YOLOv5*B. Moreover, in each of the two datasets, adding Qf to both YOLOv5B and ThrConvsB resulted in a similar performance improvement in mAP: 2.2 %-2.3 % on the Concord dataset and 0.8 %-1.6 % on the public dataset, as shown in Table 2.

Table 1
Evaluation of the detection performance of the TQGDNet in comparison to the state-of-the-art methods on both Concord and public datasets. The "CNN Model" field indicates the main deep learning architecture employed by the method.

Author (year)	CNN Model	Precision	Recall	F-1	mAP
***Concord* dataset**					
Redmon et al. (2016)	YOLOv5	0.836	0.765	0.799	0.832
Zhao et al. (2020)	Faster R-CNN	**0.896**	0.765	0.825	0.882
Follmer et al. (2022)	MTL	0.852	**0.874**	0.863	0.871
Zair et al. (2023)	SegNet-VGG16	0.855	0.855	0.855	0.843
TQGDNet (our method)	ThrConvs+ QF+GD	0.892	0.846	**0.868**	**0.918**
***OrCaScore* dataset**					
Redmon et al. (2016)	YOLOv5	0.858	0.841	0.849	0.860
Zhao et al. (2020)	Faster R-CNN	0.833	0.795	0.813	0.898
Follmer et al. (2022)	MTL	0.806	**0.917**	0.858	0.853
Zair et al. (2023)	SegNet-VGG16	0.850	0.850	0.850	0.925
TQGDNet (our method)	ThrConvs+ QF+GD	**0.890**	0.864	**0.877**	**0.948**

Table 2
Ablation of each fusion module (Qf, GD) and their combination (Qf+GD) working on two backbones: YOLOv5*B and ThrConvs*B, using both the Concord and public datasets. YOLOv5*B: Used the backbone of YOLOv5. ThrConvs*B: The backbone of TQGDNet incorporated the designed ThrConvs block into YOLOv5*B. Qf: Queen-fusion. GD: Lower layer branch of the gather-and-distribute module.

	Precision	Recall	F1	mAP
***Concord* dataset**				
ThrConvs*B	0.874	0.767	0.817	0.865
ThrConvs*B+Qf	0.845	0.805	0.825	0.888
ThrConvs*B+GD	0.886	0.809	0.846	0.906
ThrConvs*B+Qf+GD	0.892	0.846	0.868	0.918
YOLOv5*B+Qf	0.849	0.816	0.832	0.854
YOLOv5*B+GD	0.824	0.831	0.827	0.870
YOLOv5*B+Qf+GD	0.858	0.844	0.851	0.899
***OrCaScore* dataset**				
ThrConvs*B	0.767	0.972	0.897	0.895
ThrConvs*B+Qf	0.883	0.861	0.872	0.913
ThrConvs*B+GD	0.836	0.932	0.881	0.929
ThrConvs*B+Qf+GD	0.890	0.864	0.877	0.948
YOLOv5*B+Qf	0.880	0.833	0.856	0.876
YOLOv5*B+GD	0.945	0.818	0.877	0.895
YOLOv5*B+Qf+GD	0.902	0.840	0.870	0.922

We also compared YOLOv5 (in Table 1) with ThrConvs*B (in Table 2), which represents YOLOv5 with the addition of the ThrConvs module, to investigate the efficiency of the ThrConvs module. This comparison demonstrated that the ThrConvs module helped YOLOv5 increase mAP to 3.3 % and 3.5 % in the Concord and public datasets, respectively. Fig. 8 provides examples of the disparity between YOLOv5, TQGDNet, and three state-of-the-art methods employing Faster R-CNN, U-Net, and SegNet-VGG16. The fifth row in Fig. 8 reveals that YOLOv5 solely detected calcium regions exhibiting higher contrast, which had a negative impact on its performance. In contrast, ThrConvs improved its performance by effectively detecting smaller, less conspicuous calcium deposits, especially those located at the edge or near bright areas.

We further investigated the transmission of features between the network levels and layers by comparing the Qf to two other classical fusion methods, FPN and PANet. As described in Section 2.2.2 and illustrated in Fig. 5, FPN is designed as a one-way, top-down information path within the same layers, while PANet was improved by adding a bottom-up path into the FPN. Table 3 compares Qf with FPN and PANet by adding them separately to YOLOv5*B, demonstrating that Qf achieves a higher mAP than FPN by 3.4 % and 3.5 % in the Concord and public datasets, respectively.

We also assessed the impact of adding 'GD' to improve the quality of detection for small calcium instances. Since YOLOv5 was primarily utilized for detecting targets larger than calcium, adding 'GD' to YOLOv5*B resulted in a higher increase in mAP in the public dataset compared to adding it to ThrConvs*B. We also analyzed the impact of adding the combination of 'Qf' and 'GD' to enhance the final results. Interestingly, although Qf and GD separately contribute to a higher mAP in the Concord dataset based on both ThrConvs*B and YOLOv5*B, their combination (Qf+GD) benefits more in the public dataset.

4. Discussion

Our study proposed a new model to detect coronary artery calcium deposits on two datasets, including ECG-gated (high contrast) and non-gated (low contrast) CTs. Our key findings are: i) We have validated the effectiveness of TQGDNet by comparing it with state-of-the-art methods demonstrating that TQGDNet provides superior detection of small calcium deposits; ii) The ablation study demonstrated that our ThrConvs, Qf, and GD modules each contributed to overall performance improvements and, iii) TQGDNet shows superior performance between two distinct datasets, including ECG-gated on the public dataset and non-

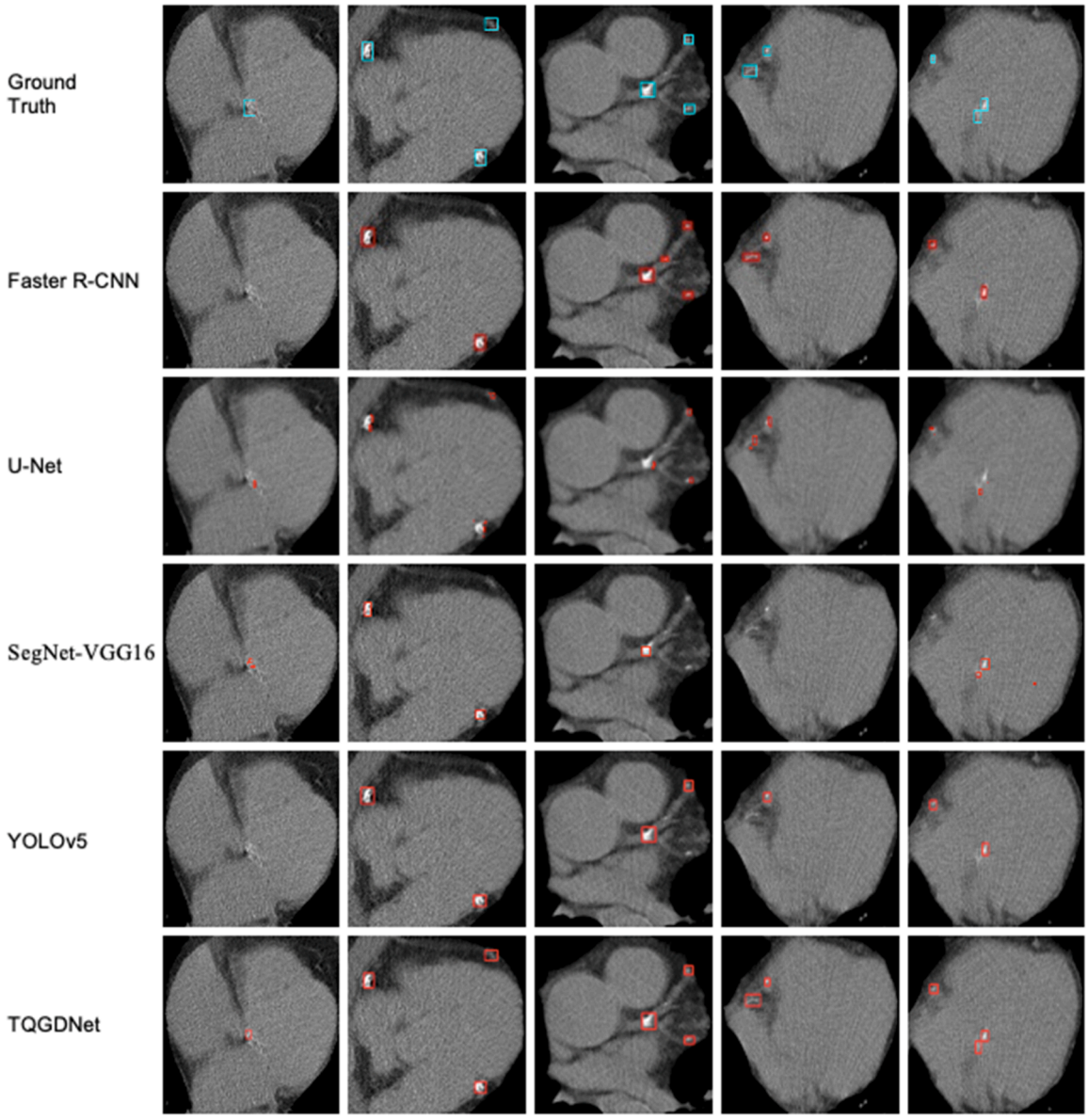


Fig. 8. Qualitative comparisons are presented between the results from YOLOv5, TQGDNet, and three state-of-the-art methods on our Concord dataset. The first row displays five input images with blue ground-truth bounding boxes for calcium deposits. The second to fourth rows show the results from state-of-the-art methods using Faster R-CNN, U-Net, and SegNet-VGG16, respectively. The sixth row presents the results from our TQGDNet. Red boxes highlight the detected calcium deposits.

Table 3
Comparison of three fusion techniques, including Qf, along with two other classical fusion techniques FPN and PANet. Each of the techniques was separately implemented on TQGDNet and tested on both our Concord cardiac CT dataset and the public cardiac CT dataset, OrCaScore.

	Precision	Recall	F1	mAP
***Concord* dataset**				
FPN	0.858	0.757	0.804	0.820
PANet	0.836	0.765	0.799	0.832
Qf	**0.849**	**0.816**	**0.832**	**0.854**
***OrCaScore* dataset**				
FPN	0.790	0.856	0.821	0.841
PANet	0.858	0.841	0.849	0.860
Qf	**0.880**	**0.833**	**0.856**	**0.876**

gated on the Concord dataset.

4.1. Comparison to the state-of-the-art

We evaluated three state-of-the-art methods optimized for the public OrCaScore dataset, utilizing the cropped heart region as input to ensure a fair and consistent basis for comparison. Among these methods, Follmer et al. (2022) provided pre-trained weights for prediction, whereas the other two were re-implemented following the methodologies detailed in their respective publications. It is important to highlight the distinction in their approaches: YOLO and Faster R-CNN are detection-based models, whereas MTL and SegNet-VGG16 are segmentation-based models. These methodological differences had a slight impact on the outcomes, as detection-based models are designed to identify and localize specific regions of interest, whereas segmentation-based models focus on providing detailed and comprehensive pixel-level classification across the entire region. This difference

in focus contributes to variations in performance metrics, reflecting the distinct objectives and capabilities of each approach.

Our method outperformed all the comparison methods. We attributed this to our method being designed to cater for non-ECG-triggered images in our Concord dataset, which is significantly different from the public OrCaScore dataset that exhibits relatively high contrast. I discussed the comparison against the state-of-the-art from three points of view: the number of images used, contrast, and small region detection. Firstly, in terms of the number of images used, Follmer et al. (2022) and Zair et al., 2023 utilized a larger number of images to train their models compared to ours. Follmer et al., 2022 released weights trained on three different ECG-gated datasets, involving two Concord datasets (DISCHARGE, CADMAN) and one public dataset (OrCaScore), totaling more than 8980 images. Even though we compared the predictions made with these pre-trained weights on the same test dataset as ours, our model, trained on only 168 images, achieved a higher mAP on both the Concord and public datasets. Moreover, the other two methods focused on operating with a dataset comprising ECG-gated images and included more images. Zhao et al. (2020) used 180 images from OrCaScore containing calcified plaques for data augmentation, expanding their dataset to 468 images. Zair et al. (2023) utilized all 3960 images from OrCaScore, including images with and without calcium deposits. Our model stands out for its incorporation of three modules designed specifically for processing non-gated images, enabling the enhancement of calcium deposits without requiring a large number of images. Secondly, detecting calcium deposits in low-contrast images poses significant challenges, particularly when these deposits have intensity values similar to surrounding structures like the pericardium or epicardial fat. The encoder-decoder structures, such as MTL (Follmer et al., 2022) and SegNet-VGG16 (Zair et al., 2023), compress the input image into low-resolution feature maps before restoring them to high-resolution. Interestingly, SegNet-VGG16 performed better on the Concord dataset compared to the public dataset. We attribute this improvement to SegNet's design, which enhances boundary delineation. This feature is particularly advantageous for the low-contrast CT images in the Concord dataset. However, this process poses a risk of loss in grabbing features between layers. Our model differs from these two methods as the ThrConvs module is designed to retain the key features and remove uncommon features by calculating between one to three convolutions. Thirdly, the detection of small regions presents challenges in extracting and retaining detailed features of these regions across different levels and layers of the CNN architecture. Follmer et al. (2022) addressed the coronary calcium scoring problem by utilizing multiple closely related calcium scoring tasks, thereby benefiting from the sharing of complementary information between these tasks. This approach included two simultaneous tasks: segmenting coronary artery regions using weak labels (labeling coronary region However, this method requires additional time and expertise to label the regions of the coronary arteries, which helps narrow down the area for detecting calcium deposits. s) and segmenting coronary artery calcifications using strong labels (labeling calcium deposits). In addition, Zhao et al. (2020) utilized optimization techniques to train the RPN efficiently, enabling the detection of small regions. However, its use of fixed scales and aspect ratios to generate the RPN, as well as the pooling operations, hinder its ability to detect small regions (Xiao et al., 2020). However, our model integrated the fusion method (via the Qf module) to preserve detailed features of small regions from the ThrConvs output by connecting multiple layers and levels, particularly in the lower layers (through the GD module) where rich small region features were located. This approach avoids the need to select sizes and aspect ratios for the RPN and mitigates information loss from pooling operations. This is especially crucial for our dataset, which primarily consists of small region representing calcium deposits.

4.2. Ablation study

The ablation study demonstrated that the three proposed modules contributed to improving the prediction performance. For ThrConvs, its design aimed to excel in detecting calcium in low-contrast scenarios, especially evident in low-dose lung CT scans without ECG-triggering, which often exhibit higher noise levels and more motion artifacts. Therefore, the incorporation of the designed ThrConvs module at the beginning of the model, aimed at primarily targeting calcium detection in low-contrast scenarios, led to improved performance (Table 2). Overall, these findings highlight the effectiveness of our proposed ThrConvs module in enhancing calcium detection in non-contrast images compared to the baseline YOLOv5.

The analysis of the Qf module revealed its capability to fuse more comprehensive features from multiple levels and layers, as demonstrated by its performance compared to state-of-the-art fusion technologies (Table 3). With the integration of the Qf module into the model, not only does it avoid the loss of information typical of the traditional FPN structure, but it also learns more comprehensive feature information (Table 1).

After receiving features from Qf, the GD module subsequently gathered and aligned those features to generate global information, which was then distributed across each level. More specifically, this process of gathering and distributing facilitated the model in learning calcium features from a broader perspective, aiding in understanding the relationship between calcium and its background and surroundings.

4.3. Limitations and future work

We identified few limitations and opportunity for future works. Firstly, the intensity of calcium deposits often closely resembles that of surrounding anatomical structures, such as epicardial fat or the pericardium. This similarity complicates the detection process, particularly in non-gated CT scans where reduced image contrast and resolution exacerbate the risk of misclassifications or missed detections. Secondly, while recent advancements in medical image processing have introduced deep learning-based multi-step approaches, these methods typically begin by segmenting anatomical structures before isolating coronary regions for calcium detection. However, such approaches rely heavily on labelled datasets and often struggle to accurately differentiate ambiguous calcium deposits from adjacent tissues, especially in both gated and non-gated CT images. Thirdly, the detection of calcium deposits in their early stages poses a significant challenge due to their minute size. These small calcifications are particularly difficult to identify in non-gated CT scans, where low visibility and poor contrast further impede their accurate detection and quantification. Lastly, the reliance on manual annotation presents notable limitations. Annotating calcium deposits is time-consuming, labor-intensive, and prone to errors or subjectivity, creating inconsistencies in the labelled datasets. Additionally, supervised learning approaches are constrained by the high cost and time required to generate large-scale annotated data, which limits their scalability and applicability. Addressing these challenges will require the development of more advanced automated solutions that reduce dependence on manual annotation and improve the detection of calcium deposits, particularly in non-gated CT imaging.

As futurework, we plan to explore 3D deep learning algorithms for tracking coronary artery flow pathways. This approach can greatly aid in predicting the coronary circulation pathway by using sequential slices, thereby replacing the current two-step (Section 3.1) limitation to narrow down the detection area. Further improvement to our model could involve designing new fusion strategies to better integrate features and widen the field of view around the calcium deposits. The traditional approaches to fuse multi-scale features between neighboring layers only accounted for features at the same level (Jia et al., 2018) or the previous level (Le et al., 2020) of the network. Although our Qf considers both the same-level and the previous-level features, it still may result in an incomplete retention of calcium features. This is because the fusion strategies impact how effectively calcium features are transmitted between levels and layers of the model, thereby influencing the subsequent

detection results.

In our study, calcium localization relied on annotations by radiologists requiring manual input. In addition, publicly available datasets for CAC detection are limited and often contain a higher number of negative samples (without calcium) compared to positive samples (with calcium). Future work could explore leveraging contrastive self-learning or self-supervised learning techniques, which have proven effective with imbalanced datasets and large numbers of unlabelled samples.

As future work, we plan to consider extending our model for calcium classification or prediction tasks to small regions in other fields of medical imaging. This includes tasks such as skin lesion classification on dermoscopic images (Bozkurt, 2022), detection of cerebral microbleeds on MRI (Luo et al., 2024), and identification of brain metastases on MRI (Ozkara et al., 2023).

As future work, we will continue to develop in the following two important areas. In clinical settings, non-gated images are often favored over gated images due to their cost-effectiveness, simpler technical requirements, and faster acquisition times, despite the presence of motion blur caused by the heart's movement during scanning. Our algorithm demonstrated the ability to use gated CT in conjunction with non-gated CT, which improved the learning of comprehensive features of CT. In future research, we will further improve the accuracy of non-gated images to match that of gated images, reducing the need for gated images.

Another area of focus is the current reliance on CNN-based approaches, which predominantly employ supervised deep learning techniques for calcium detection (and other quantification methods). While these methods have shown effectiveness, they heavily rely on the availability of annotated gated training datasets. We will conduct research on self-supervised or unsupervised learning to enable models to learn features directly from the data itself.

5. Conclusion

We introduced a new CNN model for detecting calcium deposits in both ECG-gated and non-gated CT images. Our proposed method introduced a resolution-related module (ThrConvs) designed to focus on learning features in low-contrast images. Additionally, we integrated two fusion modules to gather comprehensive feature information and focus on learning features of small calcium deposits.

CRediT authorship contribution statement

Negishi Kazuaki: Data curation, Resources. **Pathan Shahab:** Data curation, Formal analysis, Resources. **Ahn Euijoon:** Project administration, Supervision, Writing – review & editing. **Yu Christopher:** Validation, Data curation. **Wang Wei-Chien:** Writing – original draft, Visualization, Software. **Kim Jinman:** Formal analysis, Project administration, Supervision. **Naoum Christopher:** Data curation, Resources.

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.

Appendix

Cross validation

Table 4 presents the outcomes of five-fold cross-validation and their corresponding averages for both the Concord and public datasets. We note that the precision on the Concord dataset is higher than on the public dataset; however, on the public dataset, most of the recall is higher than precision. The difference in mAP between the highest and lowest results is 4.3 % on the Concord dataset and 3.1 % on the public dataset. This gap is likely caused by the significant variability in the number of images per patient, ranging from 1 to 20 slices. In practical terms, if 20 images from one patient are included in one batch size, they occupy nearly half of the test dataset. Ambiguous and difficult-to-identify images in this set are likely to negatively impact the final accuracy. Conversely, clear and easily identifiable images are expected to yield better results. Overall, five-fold groups maintained an mAP consistently close to an average of 0.896 on the Concord dataset, and an average of 0.934 on the public dataset.

Table 4
The cross-validation results for those five groups on both the Concord and public datasets

	Precision	Recall	mAP
Concord dataset			
Group1	**0.892**	**0.846**	**0.918**
Group2	0.813	0.841	0.875
Group3	0.860	0.774	0.878
Group4	0.853	0.843	0.905
Group5	0.881	0.793	0.903
Average	0.860	0.819	0.896
OrCaScore dataset			
Group1	**0.890**	**0.864**	**0.948**
Group2	0.794	0.953	0.938
Group3	0.865	0.918	0.947
Group4	0.815	0.957	0.917
Group5	0.877	0.810	0.922
Average	0.848	0.900	0.934

We present the results of the three modules used in the TQDNet using GradCAM+ + (Chattopadhyay et al., 2018) localization maps. Examples of these maps from both the Concord and public datasets are in Fig. 9, illustrating important regions from the outputs for the following model setups: YOLOv5, ThrConv, Qf, GD, and TQGDNet. With the inclusion of the ThrConv, the TQGDNet prioritized calcium deposits present in the majority of images. In the case of Qf, the activations tended to concentrate more on the calcium deposit. For GD, it exhibited a higher degree of attention towards the area around the calcium deposits. Finally, the impacts of the three modules were culminated in TQGDNet.

Fig. 9. Visualization of GradCAM++ activation maps illustrating important regions of the outputs from the use of the OrCaScore dataset and our Concord dataset, for the following model configurations: YOLOv5, ThrConv, Qf, GD, and TQGD (proposed). Red indicates higher importance, while blue indicates lower importance. Two patient studies are shown.

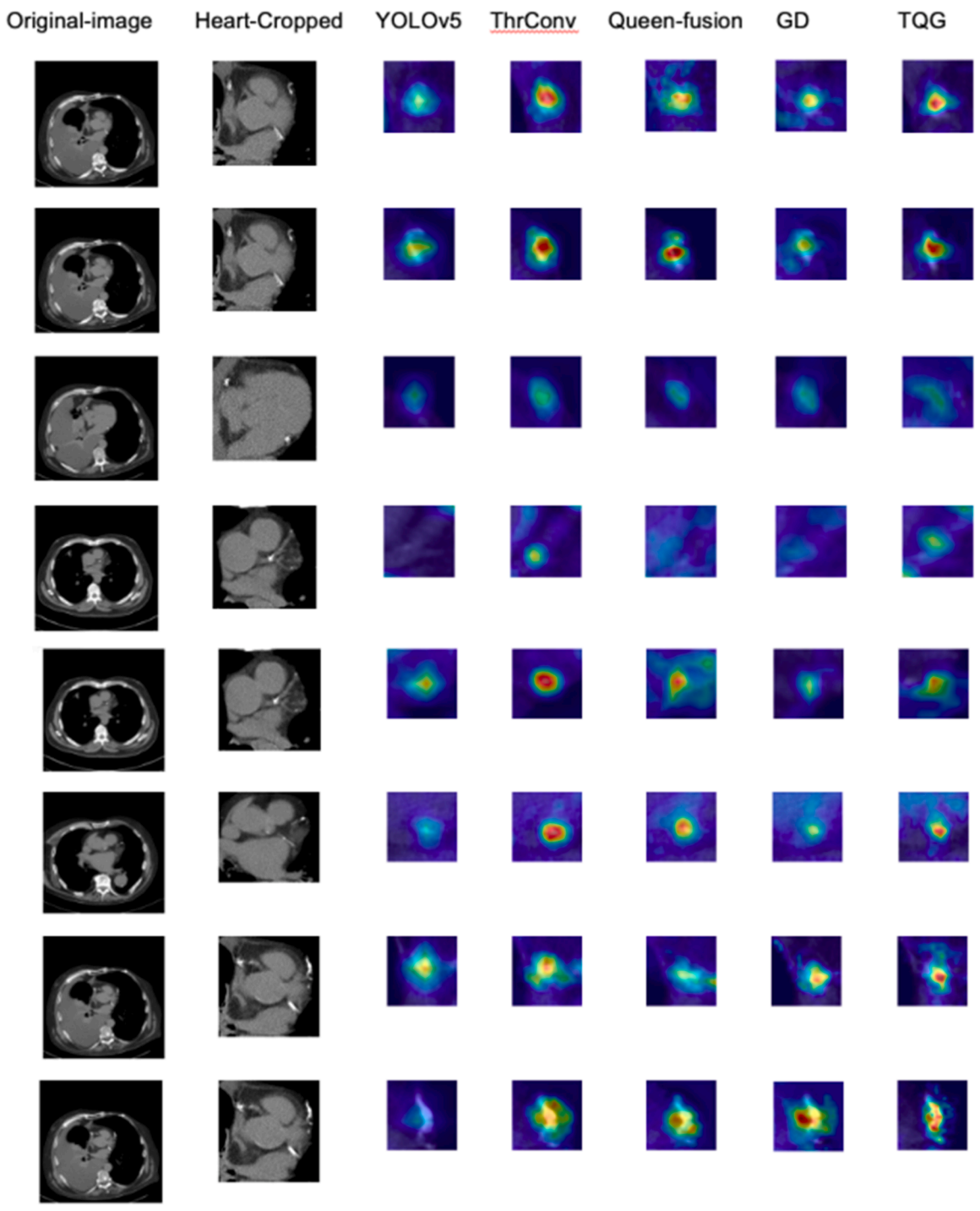


Fig. 9. (*continued*).

Data availability

This article uses two datasets: one public and one private. The author does not have permission to share the private dataset.

References

Ahmad, S., Kim, J.S., Park, D.K., Whangbo, T., 2023. Automated detection of gastric lesions in endoscopic images by leveraging attention-based yolov7. IEEE Access. https://doi.org/10.1109/ACCESS.2023.10185558.

Bhati, A., Gour, N., Khanna, P., Ojha, A., Werghi, N., 2024. An interpretable dual attention network for diabetic retinopathy grading: IDANet. Artif. Intell. Med. 149, 102782. https://doi.org/10.1016/j.artmed.2024.102782.

Bozkurt, F., 2022. Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach. Multimed. Tools Appl. https://doi.org/10.1007/s11042-022-14095-1.

Chamberlin, J., Kocher, M.R., Waltz, J., Snoddy, M., Stringer, N.F.C., Stephenson, J., Sahbaee, P., Sharma, P., Rapaka, S., Schoepf, U.J., Abadia, A.F., Sperl, J., Hoelzer, P., Mercer, M., Somayaji, N., Aquino, G., Burt, J.R., 2021. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. BMC Med. 19 (1), 55. https://doi.org/10.1186/s12916-021-01928-3.

Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-CAM++: Improved visual explanations for deep convolutional networks. arXiv preprint arXiv:1710.11063. https://arxiv.org/abs/1710.11063.

Dai, S., Han, M., Wu, Y., Gong, Y., 2007. Bilateral backprojection for single image super resolution. In: Multimedia and Expo, 2007 IEEE International Conference on, pp. 1039–1042. IEEE.

Datong, C., Minghui, L., Cheng, J., Yue, S., Dongbin, X., Yueming, L., 2019. Coronary calcium detection based on improved deep residual network in Mimics. J. Med. Syst. 43 (5), 119. https://doi.org/10.1007/s10916-019-1218-4.

Dewey, M., Rief, M., Martus, P., Kendziora, B., Feger, S., Dreger, H., Priem, S., Knebel, F., Bohm, M., Schlattmann, P., Hamm, B., Schonenberger, E., Laule, M., Zimmermann, E., (n.d.). Evaluation of computed tomography in patients with atypical angina or chest pain clinically referred for invasive coronary angiography: Randomized controlled trial. BMJ, 355.

Ding, X., Slomka, P.J., Diaz-Zamudio, M., Germano, G., Berman, D.S., Terzopoulos, D., Dey, D., 2015. Automated coronary artery calcium scoring from non-contrast CT using a patient-specific algorithm. Med. Imaging 2015: Image Process. 9413, 767–772.

Dong, W., Zhang, L., Shi, G., Wu, X., 2009. Nonlocal backprojection for adaptive image enlargement. In: Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, pp. 349–352.

Elhanashi, A., Saponara, S., Zheng, Q., 2023. Classification and Localization of Multi-Type Abnormalities on Chest X-Rays Images. IEEE Access 11, 83264–83277. https://doi.org/10.1109/ACCESS.2023.3302180.

Eng, D., Chute, C., Khandwala, N., Rajpurkar, P., Long, J., Shleifer, S., Khalaf, M.H., Sandhu, A.T., Rodriguez, F., Maron, D.J., Seyyedi, S., Marin, D., Golub, I., Budoff, M., Kitamura, F., Takahashi, M.S., Filice, R.W., Shah, R., Mongan, J., Patel, B.N., 2021. Automated coronary calcium scoring using deep learning with multicenter external validation. NPJ Digit. Med. 4 (1), 88. https://doi.org/10.1038/s41746-021-00460-1.

Follmer, B., Biavati, F., Wald, C., Stober, S., Ma, J., Dewey, M., Samek, W., 2022. Active multitask learning with uncertainty-weighted loss for coronary calcium scoring. Med. Phys. 49 (11), 7262–7277. https://doi.org/10.1002/mp.15870.

Gulati, M., Levy, P.D., Mukherjee, D., Amsterdam, E., Bhatt, D.L., Birtcher, K.K., Blankstein, R., et al., 2021. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain: executive summary: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. Circulation 144 (22), e368–e454. https://doi.org/10.1161/CIR.0000000000001030.

Haris, M., Widyanto, M.R., Nobuhara, H., 2017. First-order derivative-based super-resolution. Signal, Image Video Process. 11 (1), 1–8, 3.

Haris, M., Shakhnarovich, G., Ukita, N., 2018. Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1664-1673.

Irani, M., Peleg, S., 1991. Improving resolution by image registration. CVGIP: Graph. Models Image Process. 53 (3), 231–239.

Isgum, I., Prokop, M., Niemeijer, M., Viergever, M.A., van Ginneken, B., 2012. Automatic coronary calcium scoring in low-dose chest computed tomography. IEEE Trans. Med. Imaging 31 (12), 2322–2334. https://doi.org/10.1109/TMI.2012.2216889.

Isgum, I., Rutten, A., Prokop, M., van Ginneken, B., 2007. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. Med. Phys. 34 (4), 1450–1461. https://doi.org/10.1118/1.2710548.

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B., 2009. Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med. Imaging 28 (7), 1000–1010. https://doi.org/10.1109/TMI.2008.2011480.

Jia, S., Luo, Q., Qin, H., Qiu, J., Song, J., 2018. Path aggregation network for instance segmentation. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. https://doi.org/10.1109/CVPR.2018.00111.

Jiang, P., Li, Y., Liu, Y., Lu, N., 2023. Lymphocyte detection method based on improved YOLOv5. IEEE Access. https://doi.org/10.1109/ACCESS.2023.10374131.

Jiang, Y., Tan, Z., Wang, J., Sun, X., Lin, M., Li, H., 2022. GiraffeDet: A heavy-neck paradigm for object detection. arXiv preprint arXiv:2202.04256. ⟨https://openreview.net/pdf?id=cBu4ElJfneV⟩.

Jocher, G., 2020. YOLOv5 by Ultralytics (Version 7.0) [Computer software]. https://doi.org/10.5281/zenodo.3908559.

Kang, E., Koo, H.J., Yang, D.H., Seo, J.B., Ye, J.C., 2019. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. Med. Phys. 46 (2), 550–562. https://doi.org/10.1002/mp.13284.

Kendall, A., Gal, Y., Cipolla, R., 2017. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. arXiv preprint arXiv:1705.07115. ⟨https://arxiv.org/abs/1705.07115⟩.

Klug, M., Shemesh, J., Green, M., Mayer, A., Kerpel, A., Konen, E., Marom, E.M., 2022. A deep-learning method for the denoising of ultra-low dose chest CT in coronary artery calcium score evaluation. Clin. Radiol. 77 (7), e509–e517. https://doi.org/10.1016/j.crad.2022.03.005.

Kurkure, U., Chittajallu, D.R., Brunner, G., Le, Y.H., Kakadiaris, I.A., 2010. A supervised classification-based method for coronary calcium detection in non-contrast CT. Int. J. Cardiovasc. Imaging 26 (7), 817–828. https://doi.org/10.1007/s10554-010-9607-2.

Le, M.T., Pham, Q.V., Tran, P.Q., 2020. Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi:10.1109/CVPR.2020.00929.

Lee, S., Rim, B., Jou, S.S., Gil, H.W., Jia, X., Lee, A., Hong, M., 2021. Deep-learning-based coronary artery calcium detection from CT image. Sens. (Basel) 21 (21). https://doi.org/10.3390/s21217059.

Lessmann, N., van Ginneken, B., Zreik, M., de Jong, P.A., de Vos, B.D., Viergever, M.A., Isgum, I., 2018. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. IEEE Trans. Med. Imaging 37 (2), 615–625. https://doi.org/10.1109/TMI.2017.2769839.

Lessmann, N., Setio, I.I., de Vos, B.D., Ciompi, F., de Jong, P.A., Oudkerk, M., Mali, W.P. T., Viergever, M.A., van Ginneken, B., 2016. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. In: SPIE Medical Imaging, vol. 9785. ⟨https://ui.adsabs.harvard.edu/abs/2016SPIE.9785E.11L/abstract⟩.

Li, M., Hsu, W., Xie, X., Cong, J., Gao, W., 2020. SACNN: Self-Attention Convolutional Neural Network for Low-Dose CT Denoising With Self-Supervised Perceptual Loss Network. IEEE Trans. Med. Imaging 39 (7), 2289–2301. https://doi.org/10.1109/TMI.2020.2968472.

Li, L., Wei, M., Liu, B., Atchaneeyasakul, K., Zhou, F., Pan, Z., Kumar, S.A., Zhang, J.Y., Pu, Y., Liebeskind, D.S., Scalzo, F., 2021. Deep learning for hemorrhagic lesion detection and segmentation on brain CT images. IEEE J. Biomed. Health Inform. 25 (5), 1646–1659. https://doi.org/10.1109/JBHI.2020.3028243.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2117–2125. https://doi.org/10.1109/CVPR.2017.124.

Luo, Y., Gao, K., Fawaz, M., Wu, B., Zhong, Y., Zhou, Y., Haacke, E.M., Dai, Y., Liu, S., 2024. Automatic detection of cerebral microbleeds using susceptibility weighted imaging and artificial intelligence. Quant. Imaging Med. Surg. 14 (3), 2640–2654. https://doi.org/10.21037/qims-23-1319.

Naghavi, M., Maron, D.J., Kloner, R.A., Berman, D.S., Budoff, M., Superko, H.R., Shah, P. K., 2019. Coronary artery calcium testing: A call for universal coverage. Prev. Med. Rep. 15, 100879. https://doi.org/10.1016/j.pmedr.2019.100879.

Ozkara, B.B., Chen, M.M., Federau, C., Karabacak, M., Briere, T.M., Li, J., Wintermark, M., 2023. Deep learning for detecting brain metastases on MRI: A systematic review and meta-analysis. Cancers (Basel) 15 (2). https://doi.org/10.3390/cancers15020334.

Peng, A.W., Dudum, R., Jain, S.S., Maron, D.J., Patel, B.N., Khandwala, N., Eng, D., Chaudhari, A.S., Sandhu, A.T., Rodriguez, F., 2023. Association of coronary artery calcium detected by routine ungated CT imaging with cardiovascular outcomes. J. Am. Coll. Cardiol. 82 (12), 1192–1202. https://doi.org/10.1016/j.jacc.2023.06.040.

Purpose, OrcaScore, 2016. ⟨https://orcascore.grand-challenge.org/⟩.

Rasul, R., Budoff, M.J., Reeves, A.P., 2023. Automated coronary artery calcium segmentation in cardiac CT using combined probability map and feature analysis: a preliminary report. SPIE Med. Imaging. https://doi.org/10.1117/12.2652999.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, real-time object detection. arXiv preprint arXiv:1506.02640. ⟨https://arxiv.org/abs/1506.02640⟩.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, pp. 234-241. Springer International Publishing.

Santini, G., D. D. L., Martini, N., Valvano, G., Gori, A., Ripoli, A., Susini, C.L., Landini, L., Chiappino, D., 2017. An automatic deep learning approach for coronary artery calcium segmentation. arXiv preprint arXiv:1710.03023. ⟨https://arxiv.org/abs/1710.03023⟩.

Singh, B., Davis, L.S., 2018. An analysis of scale invariance in object detection SNIP. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3578-3587.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. ⟨https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf⟩.

Takx, R.A., de Jong, P.A., Leiner, T., Oudkerk, M., de Koning, H.J., Mol, C.P., Viergever, M.A., Isgum, I., 2014. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. PLoS One 9 (3), e91239. https://doi.org/10.1371/journal.pone.0091239.

Timofte, R., Rothe, R., Van Gool, L., 2016. Seven ways to improve example-based single image super resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1865–1873.

Tzutalin, LabelImg. GitHub repository, 2015. ⟨https://github.com/tzutalin/labelImg⟩.

van den Oever, L.B., Cornelissen, L., Vonder, M., Xia, C., van Bolhuis, J.N., Vliegenthart, R., Veldhuis, R.N.J., de Bock, G.H., Oudkerk, M., van Ooijen, P.M.A., 2020. Deep learning for automated exclusion of cardiac CT examinations negative for coronary artery calcium. Eur. J. Radiol. 129. https://doi.org/10.1016/j.ejrad.2020.109114.

van Velzen, S.G.M., Lessmann, N., Velthuis, B.K., Bank, I.E.M., van den Bongard, D., Leiner, T., de Jong, P.A., Veldhuis, W.B., Correa, A., Terry, J.G., Carr, J.J., Viergever, M.A., Verkooijen, H.M., Isgum, I., 2020. Deep learning for automatic calcium scoring in CT: Validation using multiple cardiac CT and chest CT protocols. Radiology 295 (1), 66–79. https://doi.org/10.1148/radiol.2020191621.

Vapnik, V.N., 1998. Statistical Learning Theory. Wiley, New York.

Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Wang, Y., Han, K., 2024. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. In: Advances in Neural Information Processing Systems, vol. 36.

Wang, J., Chen, K.,Yang, S., Loy, C.C., Lin, D., 2019. Region proposal by guided anchoring. arXivpreprint arXiv:1901.03278. ⟨https://arxiv.org/abs/1901.03278⟩.

World Health Organization (W.H.O.), 2021. Cardiovascular diseases (CVDs). World Health Organization. [Online]. Available: ⟨https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases⟩-(cvds).

Xiao, Y., Wang, X., Zhang, P., Meng, F., Shao, F., 2020. Object detection based on Faster R-CNN algorithm with skip pooling and fusion of contextual information. Sensors (Basel) 20 (19). https://doi.org/10.3390/s20195490.

You, C., Yang, Q., Shan, H., Gjesteby, L., Li, G., Ju, S., Zhang, Z., et al., 2018. Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. IEEE Access 6, 41839–41855. https://doi.org/10.1109/ACCESS.2018.2858196.

Yu, C., Ng, A.C.C., Ridley, L., Anjaria, M., Meier, S., Yiannikas, J., Kritharides, L., Naoum, C., 2021. Incidentally identified coronary artery calcium on non-contrast CT scan of the chest predicts major adverse cardiac events among hospital inpatients. Open Heart 8 (2). https://doi.org/10.1136/openhrt-2021-001695.

Zair, A.M., Bouzouad Cherfa, A., Cherfa, Y., Belkhamsa, N., 2023. An automated segmentation of coronary artery calcification using deep learning in specific region limitation. Med. Biol. Eng. Comput. 61 (7), 1687–1696. https://doi.org/10.1007/s11517-023-02797-z.

Zhang, W., Zhang, J., Du, X., Zhang, Y., Li, S., 2018. An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT. Computing 100, 321–339. https://doi.org/10.1007/s00607-018-0678-6.

Zhao, M., Cai, X., Liu, H., Liu, Q., 2020. Medical prior knowledge guided automatic detection of coronary arteries calcified plaque with cardiac CT. Electronics 9 (12), 2122. https://doi.org/10.3390/electronics9122122.

Zhao, Y., Wang, R.-G., Jia, W., Wang, W.-M., Gao, W., 2017. Iterative projection reconstruction for fast and efficient image upsampling. Neurocomputing 226, 200–211, 3.