A Novel 3D Recurrent R-CNN for Medical Imaging Feature Detection: A Case Study for Coronary Calcium Detection

Sarvasya, V.[1], Gotwals, R.[1] and Butler, L., PhD[2]

[1]North Carolina School of Science & Mathematics, Durham, NC, 27705, USA

[2]Cardiovascular Section, Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, 27101, USA

Abstract

Background: Artificial Intelligence (AI) has made substantial advancements in its application within the medical field due to increase in data availability. Deep learning methods have gained prominence in the use of medical imaging methods, such as magnetic resonance imaging (MRI) and computed tomography (CT) to develop AI-based prediction and detection models of multiple cardiovascular diseases. However, there still remains caveats in their accuracy and generalizability, mostly due to the quality of the images used as inputs. For example, deep neural networks trained on pristine and optimal quality images can perform poorly when tested on images that contain noise or distortions, especially when these are obtained from hospital electronic health records (EHRs)[19]. Most of these issues stem from the fact that, in convolutional neural networks (CNN), the layers are structured so that various features are found, but only those detected at the end of the model are used while the other features are treated as latent or redundant[14]. This means that the CNNs tend to struggle in detecting small, mobile, complex structures in images and, therefore, adapt poorly to variation in images[14].

Literature suggests that this can be improved by altering the network to allow stacking maps on top of each other[1]. Through each layer of a CNN, the filters contained in its nodes are applied to images to produce feature maps. By creating multiple sections of CNNs on top of each other, each layer will produce a set of stacked feature maps that can be analyzed sequentially and whose differences can be used to identify small and specific features. By using this approach, the CNN has more data it can train on (such as maps applied with sharpened kernels and sequentially subtracted images), but can process and find more features and increase the detail of examination when detecting features in an image. The overarching aim of this project is to addresses common issues in image regression within CNNs, which are distortion and variance in quality, size, and shape.

Objectives: The main aim of this study was to develop 3D CNNs which are able to detect features from images with hyper specific sensitivity, even when such images have unfiltered noise, obstruction, and distortions.

Methods: In this study, 76 CT scans were used. 53 of these images had coronary calcium indicating cardiovascular disease (cases) while 23 were controls, showing no coronary calcium. Using novel mathematical and programmatic techniques such as events with states, plane recurrence, multivariate optimizers, intersection over union analysis, and convolutional applications of linear algebra, we developed a novel 3D CNN to detect coronary calcium as well as identify the location of the deposits. When the images were passed through the network, a bounding box was placed at the detected location of coronary calcium. The network performance was assessed using the Mean-Squared Error (MSE) between

the network-generated bounding boxes and the annotated standard boxes. These outcomes were validated, through internal validation, using 76 non-gated chest CT scans to identify coronary arteries and calculate a calcium score. These scans had an 80/20 training/testing split, with 60 scans used for training and 16 used for testing. Since the scans were given in the Digital Imaging and Communications in Medicine (DICOM) format with voxel values from -1024 Hounsfield units (HU) to 4096 HU, the window leveling was applied on the mediastinum window to shrink voxel values from -110 HU to 400 HU down to 8-bit values from 0 to 255 that could be visualized as a PNG image. These were then compared to true calcium scores adjudicated by a physician.

Results: The MSE value for the testing set was, on average, less than 2, with deviance of less than 1.5 pixels for each coordinate in each 128 by 128 pixel image. The calcium score derived from the network bounding boxes showed a correlation of $R^2 = 0.921$ across all range of calcium scores.

Discussion/Conclusion: This network accurately identifies small, mobile, complex structures in non-gated chest CT scans and provides accurate calcium scores, expanding the utility of chest CT scans and providing a tool to improve diagnosis of coronary artery disease and better delegate risk-reducing cardiac disease treatments[17].

1. Introduction

Coronary calcium is calcium that builds up in coronary arteries. A substantial buildup of calcium deposits (calcium plaques) is called atherosclerosis and can narrow arteries and can be brought about by high blood pressure, obesity, diabetes and high blood cholesterol[18]. A CAC score of 0 shows that the individual has no coronary calcium buildup while an increasing calcium score (with virtually no limit) increases the risk of cardiovascular disease (CVD)[18]. Atherosclerosis is "the underlying cause of about 50% of all deaths in westernized society due to heart attacks, stroke, and peripheral arterial disease"[18]. CAC scores help clinicians assess risk of, amongst other cardiac diseases, heart attacks and stroke[18]. When CAC scores are moderate-high, clinical attention for disease prevention and/or treatment would be required. Cardiac gated computed tomography (CT) scans are conventionally used to detect CAC. Gated CT scans may require specific drugs to slow down a patient's heart rate and imaging of the heart is done only during the mid to end diastolic phase, when cardiac movement is minimized[5]. This ensures that the gated CT contains no motion distortions and clearly highlights important features, such as coronary arteries[5]. This means that gated CT scans correct coronary motion and provide accurate values for the CAC score, since it is extremely easy to detect the coronary arteries[5].

Gated CTs are much more expensive than non-gated CTs[16]. Consequently, cardiac gated CT scans are done very infrequently and only on patients who have already been diagnosed with some measure of CVD[16]. Some patients with acute heart attacks can present with sudden cardiac death[13]. Therefore, it is imperative that better predictive tools are available in a larger group of patients to detect coronary atherosclerosis. Unlike gated CTs, non-gated chest CTs are more readily available and are routinely performed for a much wider range of pathologies other than coronary artery disease[11]. For example, chest CTs are used for detecting pulmonary embolism in patients presenting to the emergency room for trouble breathing. If these non-gated CTs could be accurately evaluated for coronary artery disease, there would be less undiagnosed CVD patients who could be prescribed risk-reducing treatments and decrease the rate of instantly fatal heart attacks[11].

Non-gated CTs do not use drugs to modulate the heart rate and simply take cross-sectional images sequentially. Because images are taken during different stages of the heartbeat, these images contain motion distortion of the heart. In non-gated CTs, the coronary arteries are still in the image's field of view, but the distortion of the image precludes accurately identifying the arteries, making calculations of a CAC difficult[10]. This reduces the overall usability of these types of CTs for such cases. When non-gated CTs are used, the CAC score for these imaging scans is assigned into three categories: mild, moderate, and severe, rather than provide a numerical value[8]. This creates some issues, mostly due to variation and bias between physicians[8]. This unstandardized system also doesn't account for age, gender, or race, making it

unfeasible to use for diagnosis and treatment[7]. There is a need for automated accurate coronary artery detection and CAC quantification, which can be achieved by the new 3D CNN's[20]. This is helpful for risk prediction and identifying patients who are at high-risk of cardiac diseases and thus require risk-reducing treatments[9].

This research proposed a novel backpropagation algorithm which would allow for the training of the 3D matrices/filters that build up the convolutional layers. This algorithm would advance current scientific knowledge on how to scale up stacked networks and how to isolate specific changes throughout numerous different variables used, allowing for more specific detected features and kernel changes. This novel method is superior to the current 2D CNNs for medical image analysis because of their inability to adapt to different complex structures while 3D CNNs will be able to make accurate arterial detection possible despite motion distortion in non-gated CTs[14].

In this project, we seek to apply a novel algorithm to radiological scans to isolate image sections of increased coronary calcification and output a calcium score. More specifically this research addresses the following aims:

i) Develop a novel backpropagation neural network algorithm to allow 3D image processing
ii) Mathematically explain how this recurrent convolutional neural network (RCNN) can be integrated with this algorithm
iii) Optimize the model algorithm to be computationally and time efficient while maintaining a higher accuracy level when compared to 2D CNNs

2. Methods

The data for this experiment consisted of 76 non-gated chest CT scans containing over 8,000 images in total. 53 of the 76 scans contained coronary calcium, while 23 did not. The 76 scans were divided into 80% training dataset and 20% testing dataset for 60 training scans and 16 testing scans. Of the training dataset, there were 41 with coronary calcium and 19 without. Of the testing dataset, there were 12 with coronary calcium and 4 without.

2.1 Backpropagation Algorithm

The first part of the project is based on modification of the backpropagation algorithm for 2D CNNs. The algorithm for 2D CNN starts by characterizing the loss function as $f(x, y, z)$ and represents the result of

the forward pass, where data is passed through the biases and weights of the neural network and the deviance is calculated[3]. This is done by quantifying partial derivatives of the loss: $\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \ and \ \frac{\partial L}{\partial z}$, where L is the loss, ∂L is the change in loss, and $\partial x, \partial y, \partial z$ are the change in each of the input variables.

These partial derivatives are then used with a 2 by 2 matrix, O. O is created as a convolution between the input matrix (or image X) and the trained filter F (Eq. 1) and similar equations for the rest of the matrix (Eq. 2), shown above with the convolution between X and F.

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \quad \textbf{[Eq.1]}$$

$$O_{(i,j)} = X_{(i,j)}F_{11} + X_{(i,j+1)}F_{12} + X_{(i+1,j)}F_{21} + X_{(i+1,j+1)}F_{22} \quad \textbf{[Eq.2]}$$

Partial derivatives mentioned above $\left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial z}\right)$ are used to find the partial derivative of the loss function with respect to the matrix convolution, $\frac{\partial L}{\partial O} = \sum \frac{\left(\frac{\partial L}{\partial v}\right)}{\left(\frac{\partial O}{\partial v}\right)}$, where v represents the loss function variables $x, y, and \ z$.

$\frac{\partial O}{\partial F}$ is the change in the output image with respect to the change in the filter and is derived as shown in Eq.3:

$$\frac{\partial O}{\partial F} = \sum X_{ij} \quad \textbf{[Eq.3]}$$

Where i and j are coordinates in the X matrix.

The gradient, $\frac{\partial L}{\partial O}$, is then used as the loss gradient for the current layer and it is used to calculate the gradient shown in Eq. 4 and is applied to the previous layer by implementing Eq. 5:

$$\frac{\partial L}{\partial F} = \left(\frac{\partial O}{\partial F}\right) * \left(\frac{\partial L}{\partial O}\right) \quad \textbf{[Eq.4]}$$

$$F_f = F_i - \alpha \frac{\partial L}{\partial F} \quad \textbf{[Eq.5]}$$

While we modify the filter overall using a common matrix for the entire filter, we also calculate the specific and different gradients that are applied to each individual element i and j, separately[4].

For each i and j element in the filter, Eq. 6 is used calculate these gradients:

$$\frac{\partial L}{\partial F_{ij}} = \sum\left(\frac{\partial L}{\partial O_{a,b}} * \frac{\partial O_{a,b}}{\partial F_{ij}}\right) = \sum\left[\sum \frac{\left(\frac{\partial L}{\partial v}\right)}{\left(\frac{\partial O}{\partial v}\right)} * \sum X_{i+a-1,j+b-1}\right] \quad \textbf{[Eq.6]}$$

Where, O is the output map, L is the loss, a and b are coordinates of O, and i and j are elements in the filter.

We then add back $\alpha \frac{\partial L}{\partial F_{ij}}$ for each element (i, j) to ensure that the filter change is specific for each individual element. This allows for more precision and location of the minima.

Eq. 7 shows the final mathematical derivation of the backpropagation for regular CNNs and its extrapolation to three dimensions is the aim of the first research question of this project for a 2 by 2 filter matrix:

$$\begin{bmatrix} F_{11_f} & F_{12_f} \\ F_{21_f} & F_{22_f} \end{bmatrix} = \begin{bmatrix} F_{11_i} & F_{12_i} \\ F_{21_i} & F_{22_i} \end{bmatrix} - \alpha \begin{bmatrix} X_{11}\frac{\partial L}{\partial O_{11}} & X_{12}\frac{\partial L}{\partial O_{12}} \\ X_{21}\frac{\partial L}{\partial O_{21}} & X_{22}\frac{\partial L}{\partial O_{22}} \end{bmatrix} + \alpha \sum\left(\frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_{ij}}\right) \quad \textbf{[Eq.7]}$$

2.2 Network Structure

The second aim in this research focused on the creation of the new Recurrent RCNN structure and its functional integration. The new structure utilizes all three spatial dimensions (3D; x, y and z), as seen in the Figures 1 and 2, below. As represented in Figures 1 and 2, each individual line is a separate network that represents one section of the image set while each 'dot' as part of the broken line is a complete layer, i.e. includes x, y and z). As in Figure 1, red arrows are shown to move away from the central image while blue arrows move towards the central image. This represents the transfer of data and values, as well as how the backpropagation algorithm will 'move' backwards[15]. The green arrows in Figure 2 show how the recurrent state of this network is developed. This is where the majority of the novel 3D techniques are implemented when using medical image. Instead of just using forward propagation on this new proposed structure, we utilize lateral propagation, or sharing of information between nodes. This allows us to use a variety of image set techniques that we could not use on a single image or 2D CNNs. Examples of these techniques include IOU (Intersection over Union) analysis, parity coloring (pixels have higher values for closer numbers in the original picture), and feature overlap among potential areas for these objects (identified through an RPN).

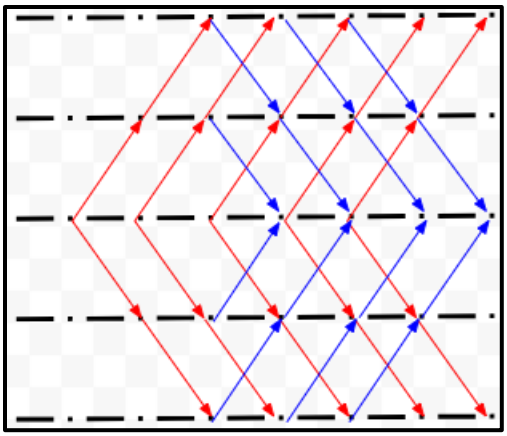


Figure 1: Modified Forward Propagation
Forward propagation technique. Black lines show movement through layers, black dots in the broken line show nodes, red lines show propagation forward and away from the center and blue lines show propagation forward and towards the center section.

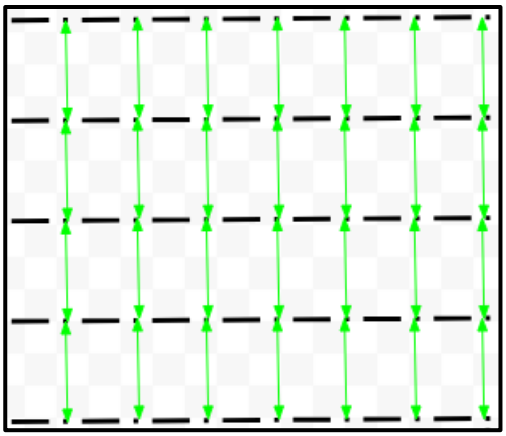


Figure 2: Introduced Lateral Propagation
Lateral Propagation technique. Black lines and black dots in the broken line represent movement through layers and nodes respectively, while green arrows show propagation laterally and communications between sections.

2.3 Modifiers

While the new backpropagation algorithm ensures that the network structure operates correctly while training on the stacked maps, the network training has not yet been optimized. One of the main issues with using 2D CNNs to detect these kinds of small, complex objects is that CNNs can only detect a singular feature or match a singular criterion for what it detects. Detecting a coronary artery in a chest CT scan requires the identification of a number of criteria to determine it as an artery, as shown below.

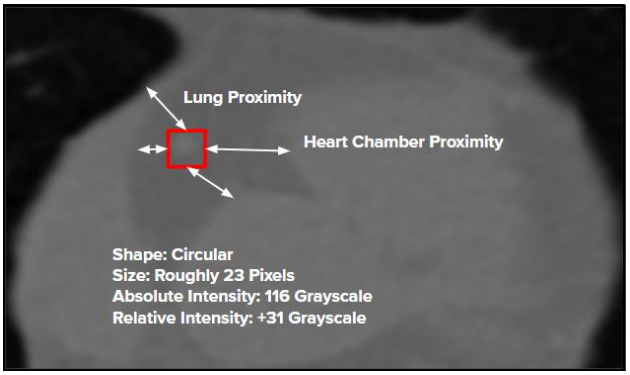


Figure 3: Arterial Conditions

To detect an artery, several considerations must be taken into account. An artery has a narrow range of proximity to the surrounding lung as well as to the main heart chambers since it exists entirely separate from the myocardial wall. The shape of the artery varies for many different slices of a single scan, and its size and location relative to the heart chambers is also liable to change. The artery's absolute intensity is needed to separate it from calcium deposits in valves, bones, and implants such as pacemakers, while its relative intensity needs to be calculated to differentiate it from contusions in the heart chambers.

2.3.1 Events with States & Plane Recurrence

Events with States is a mathematical technique used to determine how sets of numbers and probabilities affect each other. This technique deals with the entire image as a singular set for each layer. Since this method includes five layers, this can be defined as $Images = (A, B, C, D, E)$ where A, B, C, D and E are represent one section of the five full networks. As the structured algorithm moves from set to set, a coefficient of similarity, s is calculated (Eq. 8).

$$s_{A,B} = \sum ELU\left(\frac{\ln\left(\frac{A_{i,j}+B_{i,j}}{2}\right)}{\ln(255)} - \sqrt{\frac{|A_{i,j}-B_{i,j}|}{2^8}}\right) \div ij \qquad \textbf{[Eq.8]}$$

Where A and B represent two sample images, ELU is the exponential linear unit function, i and j are coordinates of their respective images.

In Eq. 8, the average of the two values of a certain pixel and take the natural logarithm (ln). The ln is then divided by the natural logarithm of 255 to allow for normalization of these values to operate in a range from 0 to 1. This is used to give higher precedence to brighter areas of the image. This is important because CT images have similar dark areas and it is important to depict similar features that need to be detected. This also helps find the ratios of white and gray areas in each image to further allow for image segmentation.

The second term within the ELU bracket is where differing pixel values are subtracted. The absolute difference between the two values are obtained and divided by 2^8. This is done to normalize such values

down to a range between 0 and 1. The square root of the absolute difference is calculated, to ensure that higher differences between images are emphasized more. The ELU function, a variant of the ReLU (Rectified Linear Unit) function was used to ensure that highly variant images have a low coefficient while still ensuring that they do not sway the calculation too much. Once ELU is applied the calculated similarity pixel values are summed over all of the pixels in the image and divided by the number of pixels to get the coefficient of similarity, introducing the recurrent element, as plane recurrence, of the network. For the forward pass of the network, a compilation of all the similarity coefficients is provided. For example, for A, a sample image, the forward pass is as shown in Eq. 9.

$$\frac{R_A+R_B s_{A,B}+R_C s_{A,C}+R_D s_{A,D}+R_E s_{A,E}}{\sum s_{A,v}} \quad \textbf{[Eq.9]}$$

Where R represents each image, and $s_{A,v}$ is the similarity coefficient between A and each of the other images.

Plane Recurrence is the other side of this coin: a connection between network sections during backpropagation and filter changing instead of forward propagation and feature map changing. While we have mostly focused on the passing forward of information between network layers, it is important to remember that there are multiple nodes in each of the five sections. Even though each node is a filter between nodes by calculating the determinant of the filter matrix.

This technique is used as a method of communication between sections to transfer information about the relative size of filters between them, alerting them to highly important and detectable features. We can ensure that the nodes on different layers are connected through a procedure using the determinants, as shown in Eq. 10.

$$F_{l,adjusted} = \frac{F_l}{\|F_l\|}\left(\sum_{k=1}^{5}(-1)^{l-k}\frac{\|F_k\|}{2^{l-k}}\right) \quad \textbf{[Eq. 10]}$$

2.3.2 Activation Function & Dense Layer Optimizer

In bounding box regression, the output layer must have four nodes: one for each of the four coordinates that define a box. For these images, the bounding box must highlight the area of the entire coronary artery so the coronary calcium can be calculated from that area. However, beyond the output layer, there is also a confidence layer, that provides another output value. This output layer determines whether the artery is detected in the image. Instead of adding this extra layer and complicating the stacked-layer network

structure, it was decided to use the ReLU activation function to serve as the output layer. For all previous layers, the ReLU function that was been used was a leaky ReLU function with negative slope of 0.1, giving the following function:

$$f(x) = \begin{cases} 0.1x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

For the last layer, the function was adjusted back down to ReLU with a maximum at 128. This means that all values below 0 and above 128 would be set to 0, while all other values would remain as x. This was done since all of the images passing through the network had dimensions of 128 by 128 pixels and therefore, so all the output values must be within this range. Output values of 0 represented non-detection of the artery instead of using the extra confidence layer.

The optimizer used on the dense layers was modified by obtaining the list of weight adjustments for that particular layer and adjusting it into a vector, $\vec{v}$. The vector then becomes $\vec{n} = \cos^{-1}\hat{v} = \cos^{-1}\left(\frac{\vec{v}}{\|v\|}\right)$ by taking the inverse cosine of each of the elements of the unit vector $\hat{v}$. Eight iterations (I) of this adjustment were done, adding 45 degrees with every iteration. This allows for the eight iterations to circle back to $360°$, giving the network a full view of all available options. Then, the cosine function was reapplied to this vector and multiplied it by its previous magnitude, giving Eq. 11.

$$\overrightarrow{v_{new}} = \cos(\vec{n} + 45°I) \qquad \textbf{[Eq.11]}$$

Where v_{new} is the new vector after each iteration. This process turned a singular change into eight different options, modified by rotating them through $length(v)$-dimensional space for weight adjustment in all different directions to improve the process of gradient descent.

2.3.3 Calcium Score Calculation

While the original images received for this project were in DICOM format, they were extracted and converted into PNG to be allow for better manipulation. This was done by setting the window level (center pixel value) and width (range of pixel intensities), you can cut only the intensities you want on the DICOM image and use those to give a more detailed image, as seen below in Figure 4.

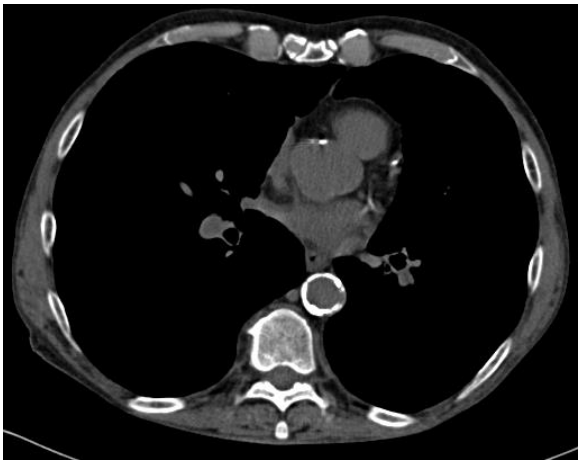


Figure 4: Window Leveling DICOM to PNG
By selecting a narrow range of values, only those values are scaled down to 8-bit PNG values, so extreme voxel values in either direction are set to either the maximum or minimum, 255 or 0 respectively, giving a clearer image highlighting only the important features.

This image not only clears away much of the noise of the previous image, but makes it easier to spot the features hidden in the image.

The Hounsfield Units (HU) of DICOM images are often used for CAC scoring. Since such values cannot be obtained from converted PNG images, a new calcium scoring algorithm must be developed based upon two values: threshold (t) and increment (i).

The usual calcium score algorithm is as follows: each lesion of calcium detected is assigned a certain scaling factor, based on its maximum intensity. If the max intensity was between 130 HU and 200 HU, it was assigned a factor of 1. Between 200 HU and 300 HU was given a factor of 2, between 300 HU and 400 HU was given a factor of 3, and anything above 400 HU was assigned a factor of 4. For HU values, it has an equivalent threshold $t = 130$ and increment $i = 100$. To find the t and i values for PNG, we had to test out all possible combinations to find the optimal one.

2.3.4 Network Validation & Software

The network was validated with 76 non-gated chest CTs, of which the first 60 were designated as the training dataset and the last 16 as the testing dataset. These images were non-contrast and originally 512 by 512 pixels. They were later shrunken down through max pooling to 128 by 128 pixels, which is the

size they were input into the network as. The images were all taken from the same source (Wake Forest Baptist Medical Center), and had standardized values for the kilovoltage peak (KVP) of 120. The field of view was also noted for each scan for accurate calcium scoring.

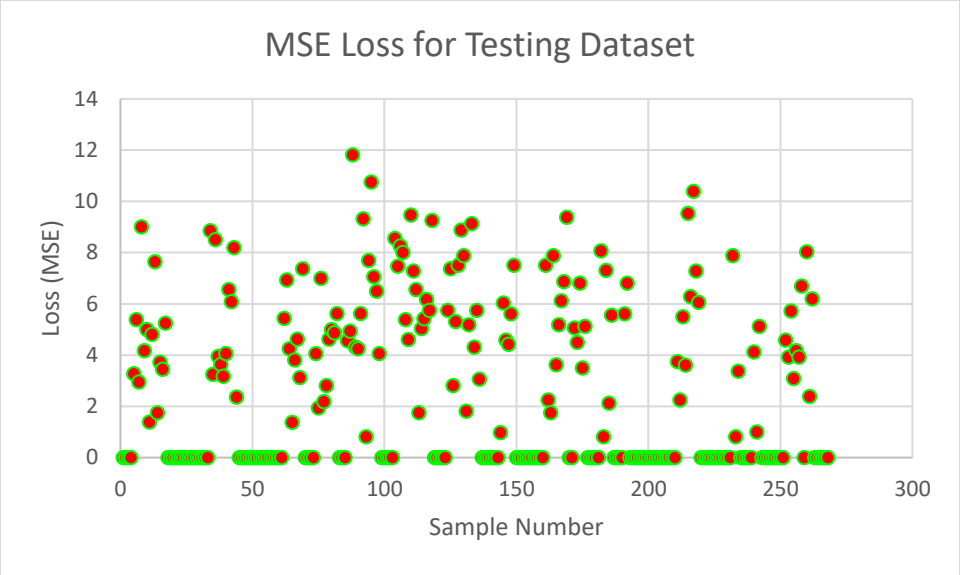
For validation of the bounding boxes, each image from each patient was taken and annotated for the location of the Left Anterior Descending (LAD) artery, the artery that was detected in this project. Four pixel values were assigned per coordinate: the Left X, Right X, Top Y, and Bottom Y, with the location of the artery being determined by these four measurements and the tuple (0, 0, 0, 0) being given if the artery was not located.

After validation with bounding boxes, the network used the calcium scoring algorithm on its generated bounding boxes to create a calcium score that was validated against a manually calculated one.

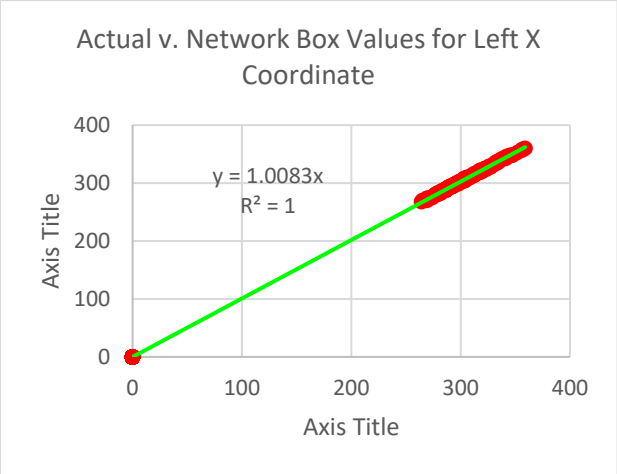
This project was undertaken using the python programming language, and the files were held on two places: an Ubuntu 20 virtual machine stored on a local disk and a personal account to the Bridges2 Pittsburgh Supercomputer. The virtual machine was used for most of the debugging, data analysis, data curation, and initial writing of the code, while these files were transferred to Bridges2 to utilize its parallel processing capabilities. The library mpi4py and its submodule MPI were used to implement parallel processing with 80 processors on the python network scripts, allowing the network to run at six seconds per image sample.

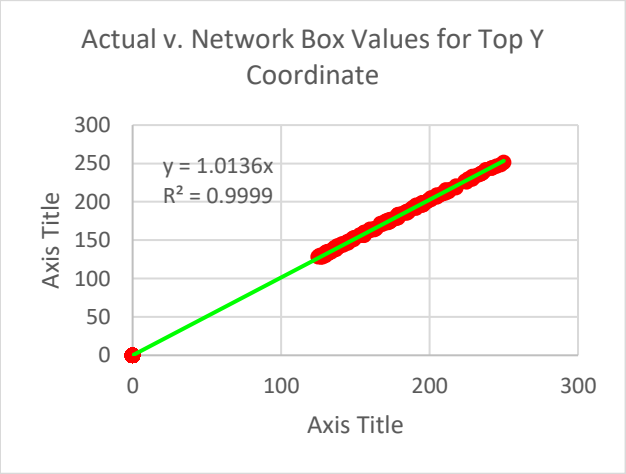
3. Results

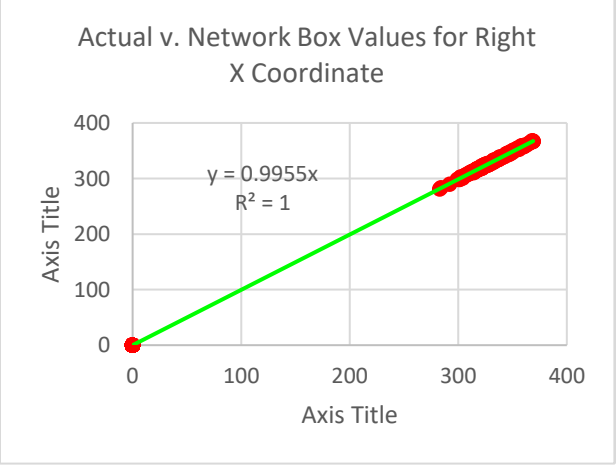
First, we can look at the bounding box accuracy through two metrics, the error calculated through Mean Squared Error (MSE) and the differences between each of the four coordinates (represented as DLTX, DLTY, DRBX, and DRBY).

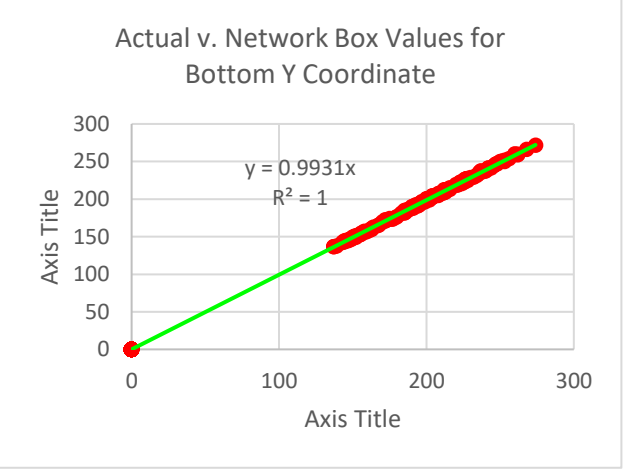


The correlation for each of the four coordinates is given below.





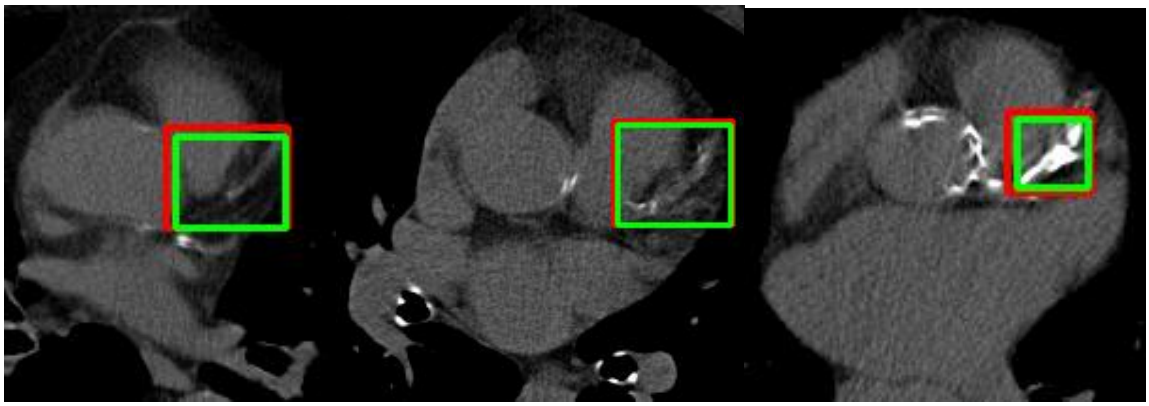




As with all neural networks, the network improves over time, getting better as training continues. Below are three tables, representing the progression of bounding box detection getting more accurate. Over time, the coordinates become more and more correlated, as noted by a p-value of 0.87 for the Left X coordinate, 0.80 for the Top Y coordinate, 0.97 for the Right X coordinate, and 0.96 for the Bottom Y coordinate. Using Pearson correlation between each network v. actual set for each of the coordinates, we see that the best correlation coefficient is reached when the Left X is 0.7% off, Top Y is 1.3% off, Right X is 0.5% off, and Bottom Y is 0.7% off, as seen by the deviance from 1 of each of the above slopes.

As we can see from these graphs, the bounding box accuracy becomes highly accurate as the network trains, finally settling at consistently within two pixels of the actual box, despite the movement of the LAD artery and differences in the orientation of the heart between images. For the full testing set, the average difference between the coordinates for the network box and actual box is three pixels. While this is highly impressive in the context of the full 512 by 512 image, the LAD artery is actually contained within a 60 by 60 pixel range. While this does diminish the effect of the network, it provides justification for why the accuracy of the bounding box regression was so much better than a normal CNN on a pristine dataset.

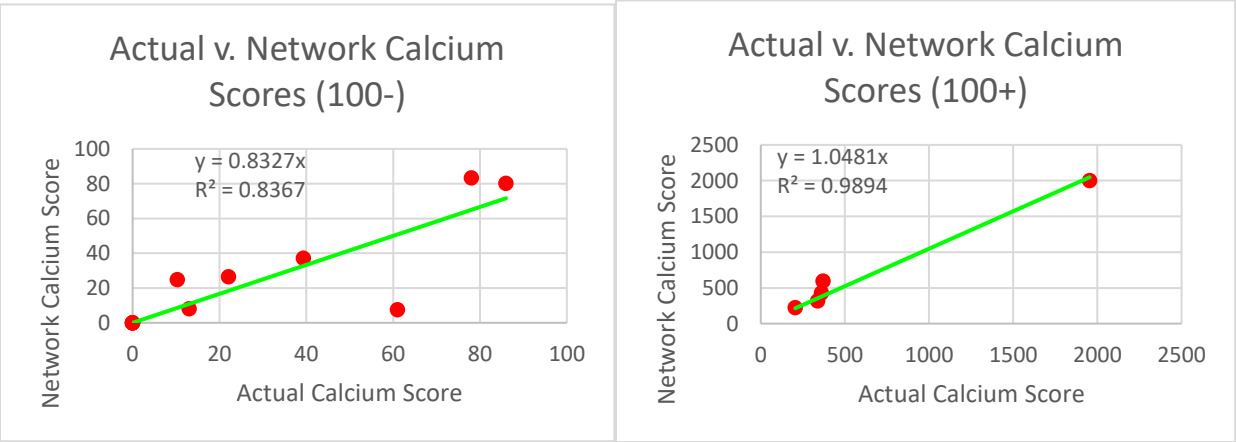
We can see a visual representation of this accuracy on the three images below, showing the detection of the artery in different circumstances. In each of these circumstances, we can see differences in the orientation of the heart, the field of view of the heart within the image, and different shapes of the artery, yet it is still able to be accurately detected.



There is another metric to determine accuracy by: Calcium scores. While the above boxes were drawn by me and could potentially contain error or deviance, the calcium scores for these scans were adjudicated by a clinician. By testing possible *threshold (t)* and *increment (i)* values as explained in the Methods section, the scale was determined to be $t = 141$ and $i = 25$ on the PNG values. We can

compare the results from the network with the calcium score calculation algorithm to get the following results.

When clinicians use calcium scores for a certain diagnosis, they group the scores into buckets to categorize the level of risk. Splitting up these calcium scores into those buckets (0, 1 to 10, 10 to 100, 100 to 400, and 400+) shows how these calculated calcium scores would actually be used, and is displayed below. The calcium scores stay in their respective buckets over 90% of the time, and even those straying outside their bucket have an average percent error to the actual score of less than 20%. We see a R^2 value of 0.921 in this correlation and a p-value between the actual scores and calculated scores of 0.86.



4. Discussion

This study attempted to identify coronary arteries using novel 3D networks and propagation algorithms. In addition to accurately identifying the coronary arteries, it automatically calculated a calcium score. Additionally, it correctly identified patients with calcium scores under 100 and above 100. The encouraging results above demonstrate the potential of such networks. These networks are capable of filtering through noisy and cluttered data and detecting objects despite changes in orientation, position, and field of view, and can even account for distortion. The innovative techniques described in this paper as well as the novel structure that account for this accuracy are byproducts of the expansion to three dimensions and the idea that a network can improve by running on several images at once. This network is capable of accurately detecting bounding boxes of small, mobile, complex objects.

While the bounding boxes provided an extremely high level of accuracy, there can be improvement in the calcium score algorithm. Most of these issues are due to having to create a new calcium score calculation algorithm since the network is running on PNG images instead of DICOM. In a clinical setting, the

network would run on the DICOM images taken straight from the scan, so the calcium score calculation would be far more accurate using Hounsfield Units.

In terms of the actual effect of this network in the current setting, it would allow for many more scans to receive calcium scores, since the chest CT scans used here are cheaper and much more common than the cardiac CT scans currently used for calcium scores. Not only does this provide another tool for clinicians to use while diagnosing a patient, but it also expands the range of patients that receive a calcium score. Since chest CTs are taken on any patient reporting cardiac, pulmonary, or abdomen-related issues or symptoms, this calcium score could reveal an underlying disease or unnoticed atherosclerosis, helping clinicians find problems fast and more accurately, enabling them to provide more effective care.

This has implications far beyond just CT scans, and even beyond medicine in general. Many of the common problems in computer vision- distortion, obstruction, variance in shape, and variance in size- can be corrected by this algorithm and through the process of learning on multiple different images. For example, a dataset of chairs would be very hard to train on for a normal network, due to differences in the chair design (eg. rolling chair v. traditional chair) and the angle at which the image was taken[6].

On this network, this issue could be corrected by taking multiple images of the same chair from many different angles. These several different images could be fed into the network at the same time, keeping the speed of training the same even as the dataset expands, and improving the quality as the network learns from many different angles, with each network in the stack attuning to a different angle or different set of characteristics[12].

Previously, the only design of three-dimensional CNN's was based on splitting up the RGB channels and having three networks in the stack: one learning on each channel[1]. The point of this was to have each network learn on a different set of criteria and pick out its own features[1]. This idea is replicated on a larger scale (with five networks in the stack) and is viable even on images without three channels, like grayscale images.

Future work has been planned for expanding the network to different arteries, adding augmented and manipulated images to make the network more robust, and introducing different analysis elements, such as Intersection over Union, to allow for better tracking features. On CT scans, there is also the possibility of calculating the risk of valve disease. This type of network would specifically be helpful in datasets with continuous data, such as a functional MRI in the medical field. Beyond medicine, this could have wide-ranging applications in real-time detection and imaging by utilizing its third dimension to analyze images over time[2].

This proof-of-concept reveals the potential of these new methods and techniques in conjunction with different network structures and training algorithms operating on multiple images at a time by tackling artery-detection and calcium scoring in chest CTs, which had not been previously done. Despite the deviance of the calcium score itself, the accuracy of the bounding box regression suggests an effective network and training system.

This network includes the products of all three research questions by mathematically creating a new optimizer and backpropagation algorithm to train on and learn from, finding a new network structure that allows for modifications between stack sections, and applying methods to improve the computational efficiency of the network such as parallel processing. While these main aims have been completed, there are still methods left untested, such as Intersection over Union analysis and other newly developed feature detection ideas.

5. Limitations

The limitations of this study include training on only one artery: the Left Anterior Descending (LAD) instead of all four available arteries, the RCA, LAD, CX, and LM. Training on these arteries would require more time to adjudicate the calcium scores for those arteries and confirm the bounding box annotations in all of the scans and slices. The second limitation is that even though this network is meant to be trained on DICOM images, it never trained on any during this process and is still equipped to run on PNG, not DICOM images. While the process of window leveling tried to ensure that the PNG image kept most of the important cardiac features in the DICOM images, testing on the DICOM images is still necessary and could pose a problem due to its wider range of pixel values.

Another limitation of the network is the size of it: four convolutional layers and three dense layers. As it stands, the network took roughly eight minutes to process a single slice, making it very computationally inefficient. To make using these seven layers feasible, I had to use the Bridges-2 supercomputer in Pittsburgh. Through my school, I gained a remote access connection to the supercomputer. I sped up the process of network training using the mpi4py library and splitting my code between 80 processors. This was done by dividing up various nodes of each layer into each processor and allowing the processors to communicate and pass information in groups of 5. By doing this, I reduced the image processing time to roughly 6.1 seconds/slice.

A third limitation of this network is the techniques used. The core idea of this research project was to aid the neural network in navigating a complex environment by giving it new techniques that it can use to

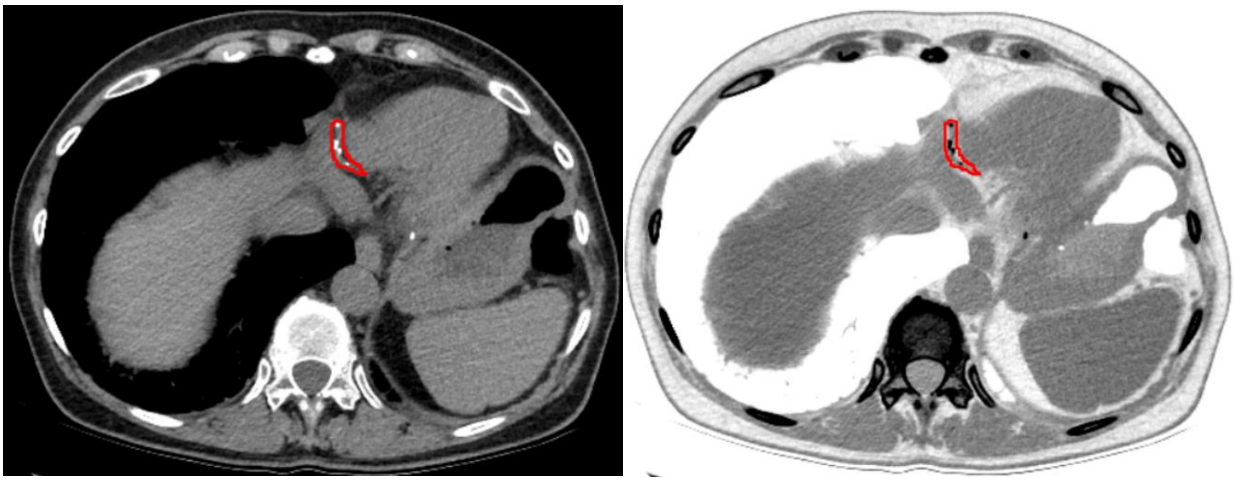
more effectively segment an image. Essentially, the aim was to give a concrete structure to a neural network's "hidden layers" by teaching the network to understand specific patterns instead of forcing it to learn by trial and error. By using these techniques, the network can be "guided" towards detecting arteries and understand it in full. The techniques used in this project do not span the full range of algorithms that can be used to optimize detection of arteries, so it will not have achieved maximum accuracy. These techniques could not be implemented due to time constraints, but they will be expanded upon in the Future Directions section.

6. Future Directions

Future directions for this project include expansion to all four coronary arteries, a larger dataset of chest CT scans, the addition of more convolutional layers, and the implementation of new techniques to further optimize the network. The following techniques would be improved methodology to further optimize the network.

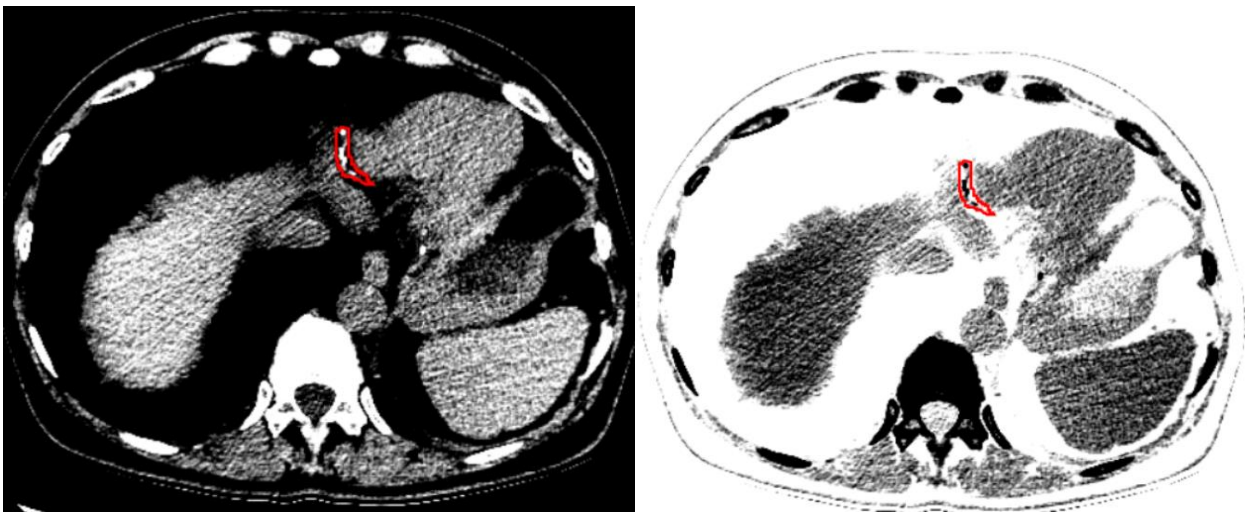
6.1 Augmented Images

One new technique to be implemented is augmented images. One of the main issues with the neural network is not being able to identify a set of criteria that the artery has to fulfill. By augmenting images through the application of convolutional kernels such as sharpen or blur or modulating the relative intensity of surrounding structures, the network is able to explore in a wider range of images and environments to become better optimized and robust.



Figures 5 & 6: Mediastinum Window, Normal and Inverse Shading

The above images are of the mediastinum window of a chest CT scan, highly similar to the images used in this project. However, we can utilize in-built DICOM tools to shade the scan slice differently and get inverse coloring, as seen in Figure 6 on the right. By using these augmented images, we can identify previously unnoticed arterial features or conditions, such as the distance between the artery and main heart chambers. While this was difficult to see before, the use of inverse shading allowed dead space to be highlighted in white, emphasizing distance between structures instead of analysis of just the artery structure itself.



Figures 7 & 8: Brain Window, Normal and Inverse Shading

The above images are of the brain window of a chest CT scan and use a much smaller range of pixels. By using a smaller range of pixels, differences between pixels are emphasized, resulting in the high variation seen in the above images. The use of different windows increases the variety and range of the dataset and can make the network more robust by helping it understand more complex features instead of superficial pixel-intensity based features. The use of inverse shading on this window also allows for new conditions to be found by differentiating the intensity of the main artery from the core heart chambers. This difference could not be spotted on a regular window, but there is a large difference on these images.

6.2 Frame Differences

Frame differences is a technique where a new image is produced by taking the absolute difference in intensity between each pixel of consecutive images, or frames. By doing this, each movement between

slices can be highly emphasized and analyzed to account for the chest CT's motion distortion. If the network were to train on these differences, it would learn how the heart moves as it processes images sequentially, and it would gain context while scrolling through the slices.

When a physician adjudicates a CT scan, most of their time is spent analyzing consecutive images to see how the artery moves and changes. By providing this capability to the neural network, it can operate contextually and is better positioned to look past noise and motion distortion to detect specific structures.

6.3 Intersection over Union Analysis

Intersection over Union Analysis is an algorithm that predicts the location of an artery given two artery locations. By calculating the number of pixels shared between two arteries over the total area encompassed by them, a fraction is produced that represents the amount of arterial motion within the image. This fraction and the location of the arteries can be turned into a "mask" applied on the next image to predict the location of the artery. This would create a "shortlist" of potential locations that the neural network has to look for and is similar to adding a Region Proposal Network (RPN) to find likely areas for arteries to be in.

6.4 Arterial Flow Modeling

This technique is not aimed at optimizing the network itself, but modifying the calcium score procedure. Currently, it is as follows:

1. Identify a lesion of calcium above the threshold
2. Find the maximum intensity of the pixels within the lesion
3. Calculate the density factor of the lesion from the maximum intensity
4. Multiply the density factor by the number of calcified pixels in the lesion
5. Multiply this quantity by the area of each pixel, as calculated from the scan field of view

While this score obviously increases for more calcium and higher intensity, it is incredibly arbitrary. To fix this, we propose a new method of calcium scoring.

$$B_{NUMER} = \sum_{i,j=0}^{h,w} \{if\ S_{i,j} \in A\} \left[max\left(\frac{S_{i,j}-t}{255-t}, 0\right)\right] \quad \textbf{[Eq. 12]}$$

$$B_{DENOM} = \sum_{i,j=0}^{h,w}\{if\ S_{i,j} \in A\}(1) \quad \textbf{[Eq. 13]}$$

$$B_{SLICE} = \left(\frac{B_{NUMER}}{B_{DENOM}}\right) \quad \textbf{[Eq. 14]}$$

$$B = \frac{\sum_{slice=0}^{len(scan)} B_{SLICE}}{len(scan)} \quad \textbf{[Eq. 15]}$$

In these equations, *i* and *j* correspond to coordinates of the image *S*, while *h* and *w* is the height and width of the image. *A* represents the set of coordinates in the artery. Eq. 12 sums the level of blockage of each pixel, with its maximum pixel intensity at 255. Eq. 13 counts the number of pixels in the artery so that they can be divided to give Eq. 14, where B_{SLICE} details the fraction of the artery that is blocked for each slice. B is calculated by averaging these values over the entire scan to find the fraction of the entire artery that is blocked. This quantity holds clinical value since it demonstrates the level of blockage in the artery and can be further used to find the level of heart stress originating from these blocked arteries and calculate the probabilities of cardiac arrest or a stroke.

6.5 Demographic Analysis & Treatment Prediction

With either the usual calcium score or the result from this new algorithm, age has to be taken into consideration[7]. For a 30-year-old, a calcium score less than 5 is highly worrying, but a score in the hundreds or even thousands is simply normal for people above 90[7]. While our data is anonymized, we do have patient ages. By correlating a patient's age with their calcium score, we can predict the acceptable range for patients of a certain age and generate a treatment plan.

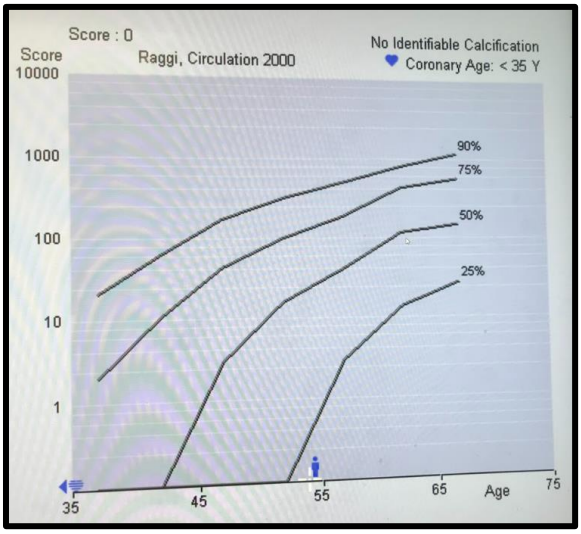


Figure 9: Calcium Score Percentiles by Age

7. Conclusion

This project created a three-dimensional R-CNN to address common issues with 2D CNNs with respect to distortion and variance in size and shape. Through techniques such as new optimizers or events with states as well as alterations in the network structure and propagation, a five-section 3D CNN was created to train on non-gated chest CT scans, which contain distorted coronary artery motion, making it an example of the improvement of 3D CNN's over 2D CNN's. The network was used to identify bounding boxes of arteries in the chest CT scan and use those boxes to calculate a coronary artery calcium score. Both the network-generated bounding boxes and calculated calcium scores showed high accuracy, making this network capable of detecting small complex structures and providing meaningful clinical info. Applications of this network include datasets with multiple continuous images, such as a functional MRI or real-time camera data, especially in a setting with high variance and movement.

The expansion an AI network to 3D analysis using stacked images would further advance medical imaging analysis knowledge by providing information on how to analyze hyper-specific differences between consecutive images to quantify both shared and isolated features. The stacked feature maps and may lead to new experiments with deriving different feature maps in the same convolutional layer. This leads to better feature isolation and a wider range of proposed features.

This new network can be applied to any kind of data that contains image sets, since the network contains stacked images as inputs. By running image sets through the network, the accuracy of feature detection could dramatically increase from the backpropagation algorithm.

References

[1]Alexandre, Luís A. "3D object recognition using convolutional neural networks with transfer learning between input channels." Intelligent Autonomous Systems 13: Proceedings of the 13th International Conference IAS-13. Springer International Publishing, 2016.

[2]B. Zhao, H. Lu, S. Chen, J. Liu and D. Wu, "Convolutional neural networks for time series classification,"
in Journal of Systems Engineering and Electronics, vol. 28, no. 1, pp. 162-169, Feb. 2017, doi: 10.21629/JSEE.2017.01.18.

[3]Cannata, Giuseppe Pio. “Backpropagation in Fully Convolutional Networks (Fcns).” Medium, Towards Data Science, 3 Feb. 2021, https://towardsdatascience.com/backpropagation-in-fully-convolutional-networks

[4]Clark, Kevin. “Deep Learning.” Stanford University Lecture Archives, Stanford University, web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/lectures/lecture5.pdf. Accessed 20 Aug. 2023.

[5]Desjardins, Benoit and Kazerooni, Ella A., ECG-Gated Cardiac CT, American Journal of Roentgenology, Volume 182, Number 4, Pages 993-1010 (2004), https://www.ajronline.org/doi/full/10.2214/ajr.182.4.1820993

[6]Dosovitskiy, Alexey, Jost Tobias Springenberg, and Thomas Brox. "Learning to generate chairs with convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[7]Greenland, Philip, et al. "Coronary calcium score and cardiovascular risk." Journal of the American College of Cardiology 72.4 (2018): 434-447.

[8]Harvey S. Hecht, Paul Cronin, Michael J. Blaha, Matthew J. Budoff, Ella A. Kazerooni, Jagat Narula, David Yankelevitz, Suhny Abbara, 2016 SCCT/STR Journal of Cardiovascular Computed Tomography, Volume 11, Issue 1, 2017, Pages 74-84, ISSN 1934-5925, https://doi.org/10.1016/j.jcct.2016.11.003.

[9]Ihdayhid, A.R., Lan, N.S.R., Williams, M. et al. Evaluation of an artificial intelligence coronary artery calcium scoring model from computed tomography. Eur Radiol 33, 321–329 (2023).

https://doi.org/10.1007/s00330-022-09028-3

[10]Išgum, I., Rutten, A., Prokop, M. and van Ginneken, B. (2007), Detection of coronary calcifications from

computed tomography scans for automated risk assessment of coronary artery disease. Med.

Phys., 34: 1450-1461. https://doi.org/10.1118/1.2710548

[11]Išgum, I., Rutten, A., Prokop, M., Staring, M., Klein, S., Pluim, J.P.W., Viergever, M.A. and van Ginneken, B.

(2010), Automated aortic calcium scoring on low-dose chest computed tomography. Med. Phys.,37: 714-723. https://doi.org/10.1118/1.3284211

[12]J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp.

4165-4170, doi: 10.1109/CAC.2017.8243510.

[13]Karam, Nicole, et al. “Incidence, Mortality, and Outcome-Predictors of Sudden Cardiac Arrest Complicating Myocardial Infarction Prior to Hospital Admission.” AHA Journals, Circulation: Cardiovascular Interventions, 4 Jan. 2019, www.ahajournals.org/doi/10.1161/CIRCINTERVENTIONS.118.007081.

[14]Laya Das, Abhishek Sivaram, Venkat Venkatasubramanian, Hidden representations in deep neural

networks: Part 2. Regression problems, Computers & Chemical Engineering, Volume 139, 2020,

106895, ISSN 0098-1354, https://doi.org/10.1016/j.compchemeng.2020.106895.

[15]Liu, Tianyi, et al. "Implementation of training convolutional neural networks." arXiv preprint (2015).

[16]Murphy, Andrew. “Cardiac Gating (CT): Radiology Reference Article.” Radiopaedia, Radiopaedia.org, 23 Mar. 2023, radiopaedia.org/articles/cardiac-gating-ct?lang=us.

[17]Nasir, Khurram, and Miguel Cainzos-Achirica. "Role of coronary artery calcium score in the primary

prevention of cardiovascular disease." Bmj 373 (2021).

[18]Pahwa, Roma, and Ishwarlal Jialal. “Atherosclerosis - Statpearls - NCBI Bookshelf.” National Library for Medicine- National Center for Biotechnology Information, National Institutes of Health, 8 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK507799/.

[19]Soukup, Daniel. "The Necessity and Pitfall of Augmentation in Deep Learning: Observations During a Case Study in Triplet Learning for Coin Images." ICPRAM. 2020.

[20]Yang, G., Chen, Y., Ning, X., Sun, Q., Shu, H. and Coatrieux, J.-L. (2016), Automatic coronary calcium

scoring using noncontrast and contrast CT images. Med. Phys., 43: 2174- 2186. https://doi.org/10.1118/1.4945045