Real-time Lesion Detection of Cardiac Coronary Artery Using Deep Neural Networks

Tianming Du[1], Xuqing Liu[1] , Honggang Zhang[1] , Bo Xu[2]

[1]Pattern Recognition and Intelligent System Lab
Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]National Center for Cardiovascular Diseases Secretary
General, CIT. Catheterization Laboratories Fu Wai
Hospital, Chinese Academy of Medical Sciences
{mercedes1993, lxq, zhhg}@bupt.edu.cn, bxu@citmd.com

Abstract: In the field of cardiac arterial interventional therapy, coronary angiography imaging provides key information to physicians for treatment strategy selection, while the lesion identification process is time-consuming and error-prone even for experienced doctors. This paper proposes a method for the automatic detection of lesion in cardiac coronary angiography based on the deep learning and convolution neural network for the very first time. We used 2925 medical images for building the model. Several lesions exist on the vessel of each image. We will regard these lesion areas as objects that are different from other background areas. We designed a model based on the convolution neural network, applying some advanced building block including CReLU, Inception and other advanced technology such as batch normalization, residual connections, skip-layer connection in our network model. After training, the network model can distinguish the difference between a lesion area and a normal vessel area (background), which can detect the location of the coronary artery lesion in real time without any manual intervention. For the stenosis lesion, the recall rate of detection achieves 0.88.

Keywords: Convolution neural network; Deep Learning; Lesion detection; Cardiac coronary artery

1 Introduction

Coronary artery disease (CAD) is one of the leading causes of morbidity and mortality worldwide [1]. As acknowledged gold standard of CAD, Digital Subtraction Angiography (DSA) plays key role for the diagnosis and analysis. Comparing to other medical images, there are less interferential biological tissues like lung or ribs in DSA. However, small vascular branches grow in abundance, which cover the artery, and many vessels overlap together influencing the distinction of the vessel distribution. Therefore, recognizing the lesion area with DSA image has become a challenging problem.

There has been a long research tradition in designing computer-aided detection systems to detect lesions automatically, improving the detection accuracy or decreasing the reading time of human experts. Numerous works take advantage of machine learning method to extract the centerline of the artery and analyze the vicinity to quantify stenosis [2][3]. Some other works segment the inner lumen to detect arterial walls to analyze the edge of the vessel [4][5]. However, the traditional method is always semi-automatic and has less ability to learn the characteristics of the CAD leading to low accuracy.

In recent years, convolutional neural network (CNN) have shown outstanding performance in medical image detection. Gülsün improved the traditional method with CNN classifier for removing extraneous paths in the detected centerlines. Zreik [7] propose a method for jointly performing both plaque and stenosis characterizations with 3D CNN model to detect the stenosis in CCTA scans automatically. Nevertheless, the process of their CNN architecture depending on pixel-level is a wasting of computing resource, resulting in slow recognition in the application.

To solve the problem above, we investigate whether a single convolutional neural network (CNN) can perform different lesion detection tasks. We propose a method for the automatic detection of lesion of coronary artery based on the non-pixel-by-pixel object detection. In this paper, we innovatively regard the different kinds of lesion as different objects in a cardiac contrast image and detect it. This detection task is quite different from traditional detection task, because the coronary artery lesion area is usually small and the patch of the lesion is quite similar to its surrounding patches (Fig. 1).

Finally, we construct an end-to-end network named CALD-Net (Coronary Artery Lesion Detection Net) for detection of coronary artery disease and tackling the above problems. Experiments show that our framework reached the real-time detection on single frame independent images, and the recall rate reached 88%, and detected the stenosis lesions that the doctor has not labeled.

The main contributions of our work are summarized as follows:

1. In our CALD-Net. we use some new popular technology that has not been verified in the lesion

978-1-5386-6067-6/18/$31.00 ©2018 IEEE

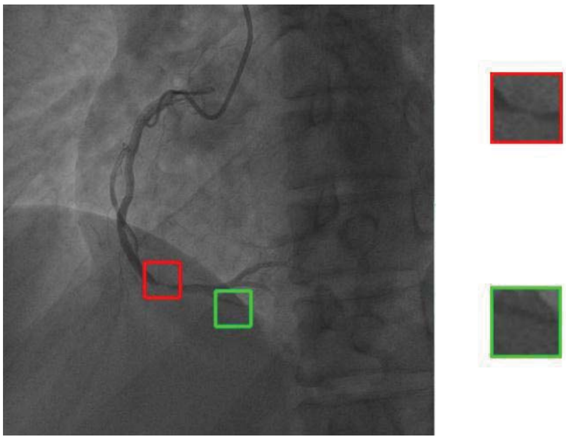

Fig. 1 The patch with the red boundary is stenosis lesion detected by our deep detection network, the patch with green boundary is the end of a healthy vessel, which is very similar to the stenosis lesion patch. Our deep detection network can distinguish the difference between these two similar patches accurately.

detection task before. A robust cardiac coronary lesion detection model has been built.

2. Extensive experiments are carried out on our cardiac coronary artery datasets. We display that our CALD-Net outperforms the previous network by large margins.

2 Methods

In our lesion detection work, all objects to be detected are lesions. We first judge whether a location is a lesion. Then we judge which kind of lesion it is. For the first step, we often generate some locations that lesion maybe exist, a good region proposal method will make good locations of lesions. We use RPN [13] to generate the region proposal. After the work of RPN, we find the location where lesion exist, then we use the feature map generated by our network to decide which kind of lesion it is. In this paper, we just detect the most common kind of lesion, stenosis, to evaluate our deep lesion detection network.

2.1 Image preprocessing

In cardiac coronary angiographic images, the peripheries of cardiac coronary are usually blurry, so morphological algorithm is used to preprocess the images. The original image is defined as x, the Top-Hat operation of morphological algorithm is defined as $Hat_{top}(x)$, and the Bottom-Hat operation is defined as $Hat_{bottom}(x)$. The Top-Hat operation enhances the edge information of the coronary arteries in the image, and the Bottom-Hat operation calculates the low value pixel of the image, highlighting the boundaries between connected arteries. The original image x plus the result of the Top-hat operation, and then subtract the result of the Bottom-Hat operation, which can enhance the contrast of the image effectively. We use the same preprocessing procedure $x_{new} = x + Hat_{top}(x) - Hat_{bottom}(x)$ for all images in our dataset.

2.2 Popular technology blocks

We use the CReLU [8] (Concatenated Rectified Linear Units) structure in first several layers of our network to reduce the computational cost of our network. Residual learning [9] method is used to overcome the training difficulty. We also use the Inception [10] and skip-layer connection [15] method, so that at multiple scales the method can extract the characteristics of the lesion to improve the accuracy of lesion detection.

2.3 Structure of CALD-Net

Combining the methods above skillfully, we proposed a new end-to-end framework for lesion detection (Fig. 2). The first block is a 7x7 convolution CReLU block named conv1_1, next block is a 3x3 max-pooling layer named pool1_1. After that there are seven 3x3 convolution blocks with CReLU activation function, named respectively conv2_1, conv2_2, conv2_3, conv3_1, conv3_2, conv3_3, conv3_4, then 8 inceptions blocks are connected with blocks above, named respectively conv4_1, conv4_2, conv4_3, conv4_4, conv5_1, conv5_2, conv5_3, conv5_4. Moreover, all convolution layers are combined with batch normalization [11] layer, channel-wise scaling and shifting, and ReLU [19] activation layer. For residual, all blocks except the first convolution block are applied. Unlike the original residual training idea, we also add residual connection onto the inception layers to stabilize the deep network architecture. Then we upscale the feature map generated by conv5_4 and downscale the feature map generated by conv3_4, and concatenate these two feature maps with the feature map generated by conv4_4. After convoluting the concatenated feature map by a 1x1 convolution layer(conv6_1), we feed the above feature map into a Faster-rcnn [12] detection network generating the region proposal, predicted the bounding- box and classification result of every ROIs. Table I shows the detailed structure of CALD-Net.

We find that the biggest size of all the stenosis lesions in our dataset is 46x149, so we should adjust the parameter of the RPN network to make it generate the proper ROIs for the detection of lesion. We set the basic feature stride size to be 16, scale to be {2, 3, 4, 5, 6} which can generate the ROIs whose area is {1024, 2304, 4096, 6400, 9216}, ratios to be {0.333, 0.5, 0.667, 1, 1.5, 2, 3} which make aspect ratios of ROIs to be {1:3, 1:2, 1:1}.

We use nearly 3000 images to train our networks. In the test phase, the input of our network is an unlabeled medical image. The output is its corresponding image that exist some bounding-box that represent the position of the lesion.

3 Experimental result

3.1 Cardiac coronary artery stenosis Dataset

We collected hundreds of Digital Subtraction

Table I The detailed structure of CALD-Net

Name	Block type	Stride	Output size	Residual
conv1_1	7x7 CReLU	2	256x256x32	
pool1_1	3x3 Pooling	2	128x128x32	
conv2_1	3x3 CReLU	1	128x128x64	√
conv2_2	3x3 CReLU	1	128x128x64	√
conv2_3	3x3 CReLU	1	128x128x64	√
conv3_1	3x3 CReLU	2	64x64x128	√
conv3_2	3x3 CReLU	1	64x64x128	√
conv3_3	3x3 CReLU	1	64x64x128	√
conv3_4	3x3 CReLU	1	64x64x128	√
conv4_1	Inception	2	32x32x256	√
conv4_2	Inception	1	32x32x256	√
conv4_3	Inception	1	32x32x256	√
conv4_4	Inception	1	32x32x256	√
conv5_1	Inception	2	16x16x384	√
conv5_2	Inception	1	16x16x384	√
conv5_3	Inception	1	16x16x384	√
conv5_4	Inception	1	16x16x384	√
upscale	4x4 Deconv	2	32x32x384	
downscale	3x3 Pooling	2	32x32x128	
concat	Concat	1	32x32x256	
conv6_1	1x1 Conv	1	32x32x512	

Angiography of coronary artery DICOM (digital Imaging Communications in Medicine) files from about 100 coronary heart disease patients. Each DICOM contains approximately 20-30 frames of coronary angiography. After selecting some frames each of which had some different kinds of lesions, such as total obstruction, stenosis, trifurcation, bifurcation, aorto ostial lesion, severe tortuosity, heavy calcification, thrombus, etc. Here we choose the stenosis lesion as our target object to detect. In each frame of the image, with the help of professional doctors we marked 1-2 stenosis, the candidate box for each lesion was represented by four coordinate values $\{x_1, y_1, x_2, y_2\}$, where (x_1, y_1) denotes the coordinate of the top-left corner of bounding box, (x_2, y_2) denotes the coordinate of the lower-right corner of bounding box. Finally, we generated 5248 candidate boxes from 2925 frames. We use these data to train our network. Fig. 3 shows some data and its ground-truth labelled by doctors.

We did not use the position information (The exposure angle of the X-ray relative to the heart. The different cardiac coronary vessel in the images that have the same position usually have the similar structure) of these pictures and other information about the patients, we used only the graphics information of these cardiac images to train our networks.

3.2 Training and evaluation

For our experiment, we divide our 2925 images into three parts: 2364 images for training, 260 images for validation, and 301 images for testing. From 2624 training and validation, 5248 lesion ground-truth bounding boxes are generated to be fed into our network. All images' size is 512x512.We set 15 scales for a training image. The size of the edge of an image maybe one value in the set {416, 448, 480, 512, 544, 576, 608, 640, 672, 704, 736, 768, 800, 832, 864}, each size corresponds to a scale.

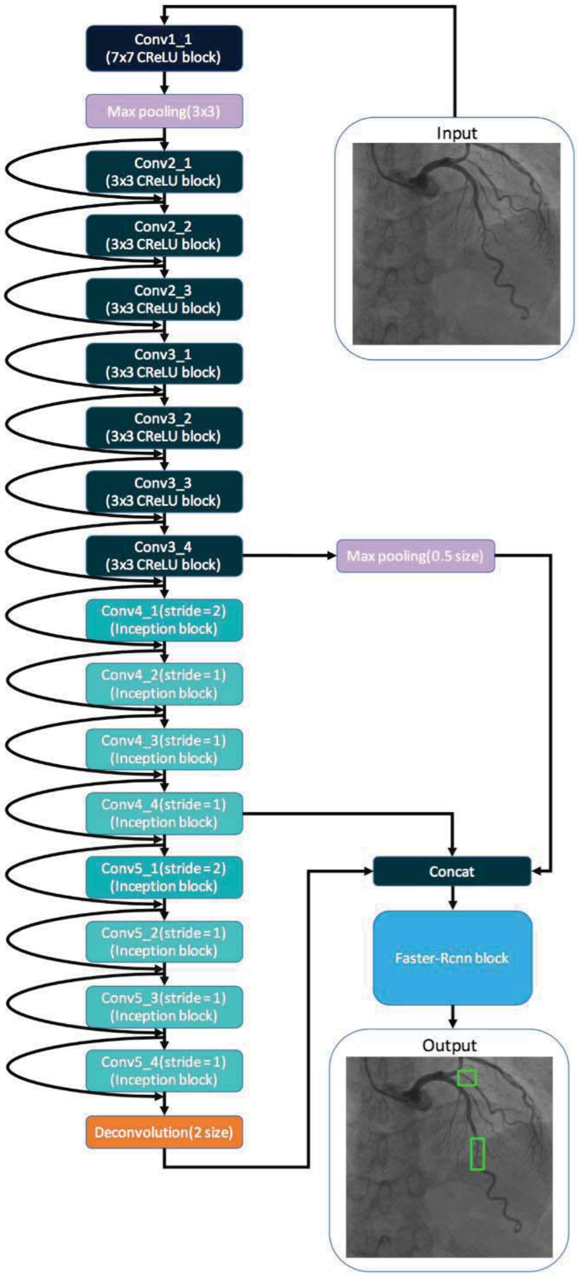

Fig. 2 The structure of CALD-Net. The curve lines in the figure is residual connection that ignore the Eltwise layer for the convenience of display.

Since our Cardiac Coronary Artery Dataset is small compared to the general image classification datasets, so before training our net using our medical images, we pre-train the model by VOC 2007 dataset using the transfer-learning scheme. In VOC 2007 detection task, the number of detection class is 21 including the background. We first take the model which is pre-train on VOC 2007 detection task and replace its input layer and output layer by a new detection class number of 2 (background and stenosis). Then, we fine-tune its parameters on our Cardiac coronary artery dataset.

When training the network, each image is resized to a random scale before being fed into the network. The

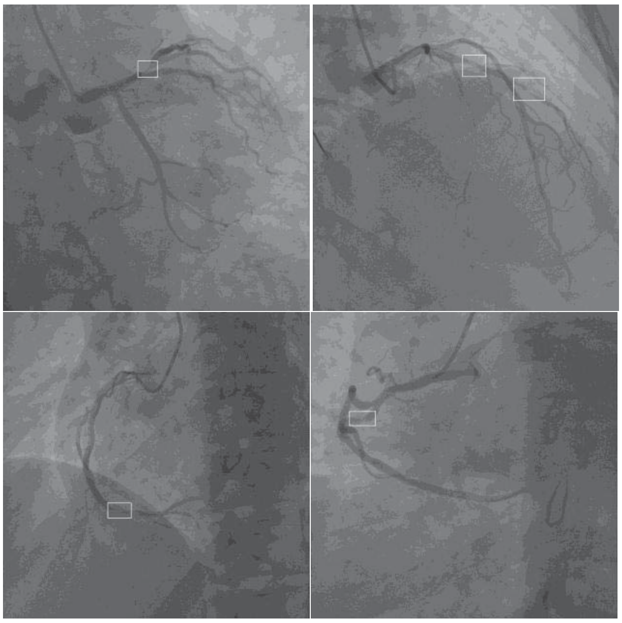

Fig. 3 Some images in our Cardiac coronary artery stenosis dataset. The white bounding-box is the ground-truth labelled by professional doctors.

threshold of positive sample for RPN is set to 0.7; the one of negative sample is set to 0.3. We use a dynamic strategy to adjust the learning rate base on plateau detection [9]. The learning rate of our network is set to 0.1 in the beginning, and decreased by a factor of $\sqrt{10} \approx 3.1623$ whenever a plateau is detected. We train our network for 200K iterations.

For our dataset evaluation, all evaluations were done on Intel Xeon CPU E5-2630 v4 CPU and NVIDIA GTX 1080 Ti GPU.

3.3 Result of CALD-Net

Table II shows the result of our CALD-Net compared with other deep networks such as ZF-Net [16], VGG [17], ResNet [9]. For ZF-Net and VGG, we connect its first five convolution layers (including the pooling layer and norm layer behind each of the convolution layer, except the 5th pooling layer) to a Faster-rcnn block to evaluate. We take the time cost, mAP and recall rate as our standard of evaluation. Recall rate refers to a ratio of "true positive sample of stenosis lesion" boxes among the proposals, considering a box as "true positive sample" if the intersection-over-union (IoU) score with its maximally-overlapped ground-truth box is more than a threshold α. The threshold α for Table I is 0.5. Table III shows the different recall rates for different α={0.3,0.5,0.7}.

Different doctors labeled our data in different time, so the standards for stenosis lesion are not the same for all data. However our CALD-Net still achieve great lesion detection results. Fig. 4 shows some results of our CALD-Net when α=0.5. It is notable that our CALD-Net

Table II Comparisons between our CALD-Net and some other state-of-the-art in our Cardiac coronary artery stenosis Dataset

Model	Recall	mAP	Time(ms)	FPS
CALD-Net	0.88	0.46	52	19.2
ZF-Net +Faster Rcnn	0.41	0.20	38	26.3
VGG + Faster Rcnn	0.50	0.26	113	8.8
ResNet50 + Faster Rcnn	0.62	0.32	940	1.1

Table III Different recall rates for different α

Model	Recall
CALD-Net (with $\alpha = 0.3$)	0.94
CALD-Net (with $\alpha = 0.5$)	0.88
CALD-Net (with $\alpha = 0.7$)	0.61

can detect the stenosis lesions that the doctor has not labeled which decrease the detection precision rate, so ours mAP is not very high, but our CALD-Net has the ability to detect the lesion precisely.

4 Conclusions

For the problem that large calculation amount of pixel-wise detection task and the difficulty in detection for the small coronary artery lesion area. We propose a method for the automatic detection of lesion of coronary artery based on the deep learning and convolution neural network. We build a new end-to-end deep network CALD-Net. After our evaluation, the CALD-Net can detect the stenosis lesion precisely, and recall rate achieve 88%, detecting the stenosis lesion that doctors leave out. In this paper, we only detect one kind of lesion for the convenience of labelling the dataset. In fact, if we label different kinds of lesions, such as total obstruction, stenosis, trifurcation, bifurcation, aorto ostial lesion, severe tortuosity, heavy calcification, thrombus, the precision rate of detecting the one specific lesion will increase, because the more kinds of lesions fed into the network, the more representative feature for one specific lesion to be generated. The tendency above has been proved in our other experiments. In addition, we are doing this work for our cardiac coronary artery stenosis dataset by adding label types of other kinds of lesion to higher precision rate of detecting the lesion in cardiac coronary artery. Furthermore, we can use the CALD-Net to detect different kinds of lesions, and use other technology to distinguish different kinds of coronary artery in order to calculate the SYNTAX [18] score without human intervention. This is also our future work.

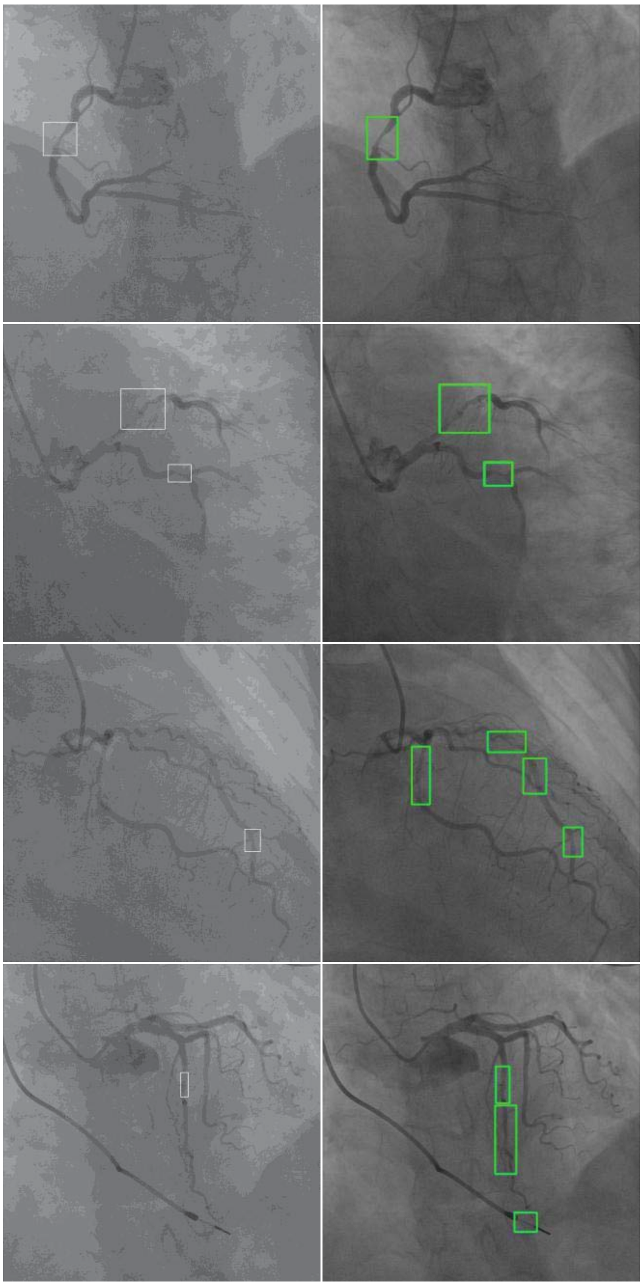

Fig. 4 The white bounding boxes in left side is the ground-truth labeled by professional doctors. The green bounding box in the right side is the detection result of our CALD-Net. Obviously, we detect some stenosis lesions that doctors leave out.

References

[1] Scarborough PB, Wickramasinghe K, et al., Coronary heart disease statistics, British Heart Foundation, 2010.

[2] Mittal S, Zheng Y, Georgescu B, et al. Fast automatic detection of calcified coronary lesions in 3D cardiac CT images[C]//International Workshop on Machine Learning in Medical Imaging. Springer, Berlin, Heidelberg, 2010: 1-9.

[3] Zuluaga M A, Magnin I E, Hoyos M H, et al. Automatic detection of abnormal vascular cross-sections based on density level detection and support vector machines[J]. International journal of computer assisted radiology and surgery, 2011, 6(2): 163-174.

[4] Wang C, Moreno R, Smedby Ö. Vessel segmentation using implicit model-guided level sets[C]//MICCAI Workshop" 3D Cardiovascular Imaging: a MICCAI segmentation Challenge", Nice France, 1st of October 2012. 2012.

[5] Broersen A, Kitslaar P, Frenay M, et al. FrenchCoast: fast, robust extraction for the nice challenge on coronary artery segmentation of the tree[C]//Proc. of MICCAI Workshop” 3D Cardiovascular Imaging: a MICCAI segmentation Challenge. 2012.

[6] Gülsün M A, Funka-Lea G, Sharma P, et al. Coronary centerline extraction via optimal flow paths and CNN path pruning[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2016: 317-325.

[7] Zreik M, van Hamersvelt R W, Wolterink J M, et al. Automatic Detection and Characterization of Coronary Artery Plaque and Stenosis using a Recurrent Convolutional Neural Network in Coronary CT Angiography[J]. arXiv preprint arXiv:1804.04360, 2018.

[8] Shang, W., Sohn, K., Almeida, D., & Lee, H. (2016). Understanding and improving convolutional neural networks via concatenated rectified linear units. International Conference on Machine Learning, pages 2217-2225.

[9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Computer Vision and Pattern Recognition. IEEE, pages 770-778.

[10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D., et al. (2015). Going deeper with convolutions. Computer Vision and Pattern Recognition, pages 1-9

[11] Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning, pages 448-456.

[12] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell, 39(6), pages 1137-1149.

[13] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. International Conference on Artificial Intelligence and Statistics, pages 315-323.

[14] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., & Fu, C. Y., et al. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision, pages 21-37.

[15] Kong, T., Yao, A., Chen, Y., & Sun, F. (2016). Hypernet: towards accurate region proposal generation and joint object detection. Conference on Computer Vision and Pattern Recognition, pages 845-853.

[16] Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional networks. European Conference on Computer Vision, pages 818-833.

[17] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Computer Science.

[18] Sianos, G., Morel, M. A., Kappetein, A. P., Morice, M. C., Colombo, A., & Dawkins, K., et al. (2005). The syntax score: an angiographic tool grading the complexity of coronary artery disease. Eurointervention Journal of Europcr in Collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology, 1(2), pages 219.

[19] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. International Conference on Artificial Intelligence and Statistics, pages 315-323.