

Comparative Study of Deep Learning Models for Automatic Coronary Stenosis Detection in X-ray Angiography

Viacheslav Danilov^{1,3}[0000-0002-1413-1381], Olga Gerget¹[0000-0002-6242-9502],
Kirill Klyshnikov²[0000-0003-3211-1250], Evgeny Ovcharenko²[0000-0001-7477-3979],
and Alejandro Frangi³[0000-0002-2675-528X]

¹ Tomsk Polytechnic University, Tomsk, 634050, Russia

viacheslav.v.danilov@gmail.com, gerget@tpu.ru

² Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, 650002, Russia

{klyshnikovk, ov.eugene}@gmail.com

³ University of Leeds, Leeds, LS2 9JT, United Kingdom

{V.Danilov, A.Frangi}@leeds.ac.uk

Abstract. The article explores the application of machine learning approach to detect both single-vessel and multivessel coronary artery disease from X-ray angiography. Since the interpretation of coronary angiography images requires interventional cardiologists to have considerable training, our study is aimed at analysing, training, and assessing the potential of the existing object detectors for classifying and detecting coronary artery stenosis using angiographic imaging series. 100 patients who underwent coronary angiography at the Research Institute for Complex Issues of Cardiovascular Diseases were retrospectively enrolled in the study. To automate the medical data analysis, we examined and compared three models (SSD MobileNet V1, Faster-RCNN ResNet-50 V1, Faster-RCNN NASNet) with various architecture, network complexity, and a number of weights. To compare developed deep learning models, we used the mean Average Precision (mAP) metric, training time, and inference time. Testing results show that the training/inference time is directly proportional to the model complexity. Thus, Faster-RCNN NASNet demonstrates the slowest inference time. Its mean inference time per one image made up 880 ms. In terms of accuracy, Faster-RCNN ResNet-50 V1 demonstrates the highest prediction accuracy. This model has reached the mAP metric of 0.92 on the validation dataset. SSD MobileNet V1 has demonstrated the best inference time with the inference rate of 23 frames per second.

Keywords: Stenosis Detection, X-ray Angiography, Deep Learning, Transfer Learning

1 Introduction

Coronary artery disease (CAD) is the leading cause of mortality worldwide [1]. CAD is commonly caused by atherosclerotic plaques encroaching the coronary artery lu-

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

men and resulting in its narrowing or complete blockage. To date, invasive coronary angiography is the gold standard for diagnosing coronary artery stenosis using X-ray visualisation of a radiopaque agent. Therefore, analysis and interpretation of coronary angiography data play an important role in the accurate diagnosis of coronary artery stenosis. The severity of stenosis and the SYNTAX score are used for selecting either minimally invasive extravascular surgery or invasive intervention.

Despite recent advances in diagnostic tools and algorithms, capable of detecting the location of coronary artery stenosis (82 - 95%) and classifying it (80 - 97%) [2–8] there are significant limitations, necessitating further studies. Major drawbacks include poor scalability and flexibility of preprocessing algorithms that require fine-tuning. Most detection algorithms use the cascading principle, prone to the accumulation of errors. Therefore, our study is aimed at developing, training, and assessing several neural networks to determine coronary artery stenosis with the highest predicting accuracy on original angiographic imaging series.

2 Source data

One hundred patients who underwent coronary artery angiography using angiography systems Coroscop (Siemens) and Innova (GE Healthcare) at the Research Institute for Complex Issues of Cardiovascular Diseases (Kemerovo, Russia) were retrospectively enrolled in the study. Patients with multivessel CAD were excluded from the study. Angiographic imaging series of the radiopaque overlaid coronary arteries with stenotic segments were selected by an interventional cardiologist. Thus, 8325 input images in grayscale (one channel) of 512×512 pixels to 1000×1000 pixels were ultimately included for further study. Of them, 7492 (90%) images were used for training, and 833 (10%) images were used for validation. Data were labelled using a free, open-source version of SaaS (Software as a Service) solution – LabelBox. Typical data labelling of the source images is shown in Fig. 1.

To analyse the source dataset, we estimated the size of the stenotic region computing the area of the bounding box. Similarly to the Common Objects in Context (COCO) dataset, we divided objects by their area into three types: small ($area < 32^2$), medium ($32^2 \leq area \leq 96^2$), and large ($area > 96^2$) objects. A total of 2509 small objects (30%), 5704 medium objects (69%), and 111 large objects (1%) were obtained in the input data. Considering the unbalanced distribution of classes in much training data, we suppose that the models may perform poorer on larger objects than on small and medium ones.

To determine the stenosis location accurately, we evaluated the distribution of the stenosis coordinates along the vessel in the input images. The coordinates of the centre point of the bounding box around the stenotic lesion were normalised and assessed. Based on this assessment, a distribution map of the coordinates of the stenosis centres was generated and is shown in Fig. 2. The distribution of the coordinates highlights two centres with relative coordinates $(0.50; 0.20)$ and $(0.27; 0.27)$ along the stenotic vessel segment. The coordinates of the centres are evenly distributed without explicit statistical outliers.

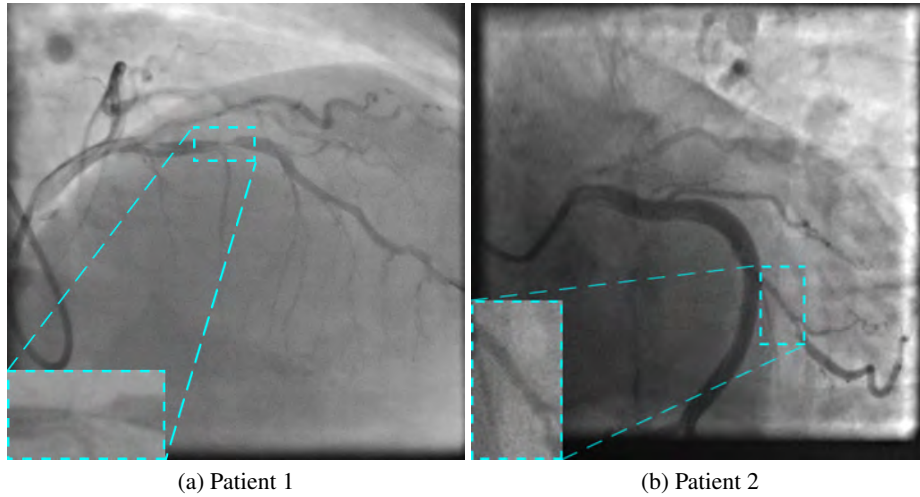


Fig. 1. Data labeling of the source images.

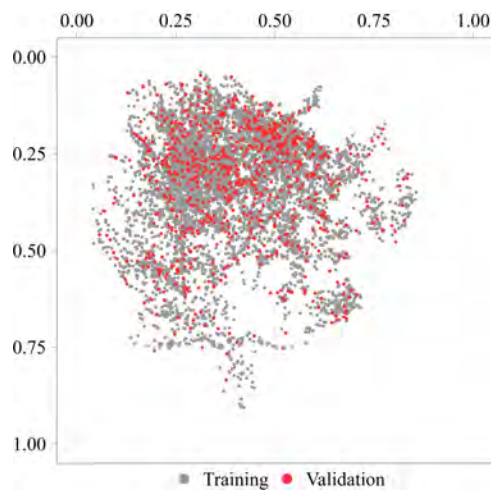


Fig. 2. Distribution map of stenosis.

3 Methods

3.1 Models description

We applied machine learning algorithms to detect coronary artery stenosis on the CAG imaging series. Machine learning has shown beneficial potential in computer vision and image processing. We used SSD [9] and Faster-RCNN [10] object detectors from the Tensorflow Detection Model Zoo [11] based on such models as MobileNet [12, 13], ResNet [14, 15] and NASNet [16, 17]. Three models with various architectures, network

complexity, and a number of weights were selected. The lightweight SSD MobileNet V1 SSD detector enabling real real-time data processing was chosen as the reference model. Faster-RCNN NASNet, with over 80 million weights, was the most complex model selected for the study. A brief description of the models is presented in Table 1. Characteristics of neural networks, including mAP, are reported based on their training on the COCO dataset.

Table 1. Brief characteristics of the selected pre-trained models.

Model	Inference time, ms	mAP	Weights, mln	Size, Mb
SSD MobileNet V1	56	32	4.2	44
Faster-RCNN ResNet-50 V1	89	30	25.6	114
Faster-RCNN NASNet	540	–	88.9	416

3.2 Models training

When training neural network models, their base configuration is similar to that used to train on the COCO dataset. For the unambiguous comparison of the selected models, the total number of training steps was set to 100 equal to 100000 iterations of learning. Regarding the loss functions, Weighted Smooth L1 loss (see equation 3 in [18]) was the localisation loss, and Weighted Focal Loss [19] was the classification loss. It should be noted that the SSD-based model was trained using the Cosine decay with the warm-up. This technique allowed gradually decreasing the learning rate (LR) depending on the learning step. To train the networks, we used P2 (Nvidia Tesla K80 12 Gb, 1.87 TFLOPS) and P3 instances (Nvidia Tesla V100 16 Gb, 7.8 TFLOPS) from Amazon Web Services. Table 2 summarises the main characteristics of the model training.

Serial changes in precision were tested on the validation set during the training process. The mAP metric, as the metric of interest, with a predefined threshold value for Intersection over Union equal to 0.5 (mAP@0.5) was used. Fig. 3 shows smooth changes in the mAP on the validation set during the training process. As seen, all models converge to a specific value of the mAP asymptotic accuracy.

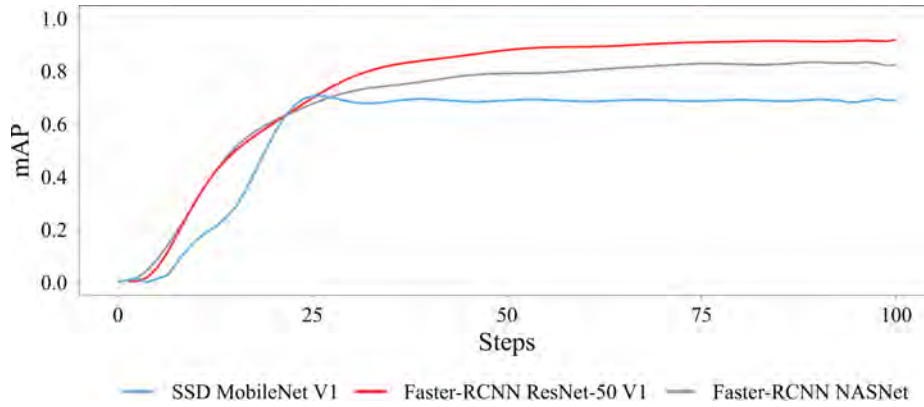
4 Results

4.1 Comparative analysis

Table 3 reports the results of the comparative study of the three selected neural network models. As we mentioned before, metrics of SSD MobileNet V1 were used as the reference. Faster-RCNN ResNet-50 V1 and Faster-RCNN NASNet models were compared

Table 2. Model training settings.

Model	Input size	Augmentation	Batch size	LR type	LR
SSD MobileNet V1	640x640x3	Random horizontal flip Random crop	4	Cosine decay	0.04
Faster-RCNN ResNet-50 V1	600x600x3	Random horizontal flip	2	Constant LR	0.0003
Faster-RCNN NASNet	1200x1200x3	Random horizontal flip	1	Constant LR	0.0003

**Fig. 3.** Changes in the mAP metric during the training process.

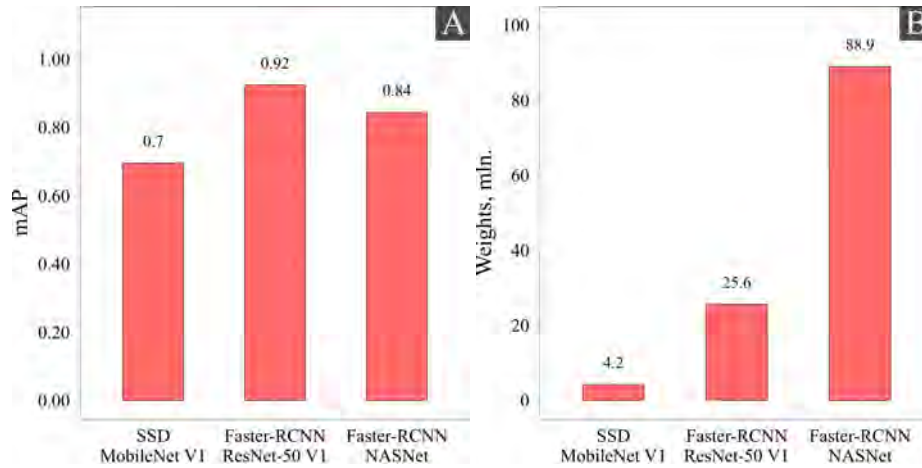
with it. Fig. 4 and 5 report the basic metrics of the model performance (mAP, training time, and inference time). The graphs are sorted in a way to present the models with superior metrics first.

The inference time was estimated using P3 instance (Nvidia Tesla V100 16 Gb, 7.8 TFLOPS) of Amazon Web Services. Based on the obtained results, the inference time directly depends on the complexity of the model and the total number of its weights. Thus, Faster-RCNN NASNet was the slowest in predictions. Its mean processing time per one image was 880 milliseconds. The model based on the MobileNet backbone was the fastest one with the inference time per one image of 43 milliseconds. Thus, it may be used for predicting the location of stenosis in real-time.

Our results suggest that Faster-RCNN ResNet-50 V1 is the most accurate model. The mean Average Precision of this model on the validation set is 0.92 with the inference time of 98 milliseconds per image (≈ 10 frames per second). The fastest and relatively lightweight SSD MobileNet V1 model has the mean Average Precision of 0.70 with the inference time of 43 milliseconds per image (≈ 23 frames per second). Faster-RCNN NASNet has over a 3-fold advantage in the number of weights compared

Table 3. Comparative study of the selected models

Model	Weights, mln		Training time, hours		Inference time, ms		mAP@0.5	
	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.
SSD MobileNet V1	4.2	1.0x	16	1.0x	43	1.0x	0.70	1.00x
Faster-RCNN ResNet-50 V1	25.6	6.0x	28	1.8x	98	2.3x	0.92	1.33x
Faster-RCNN NASNet	88.9	21.0x	147	9.5x	880	20.4x	0.84	1.22x

**Fig. 4.** Comparison of the mAP metric (a) and weights (b) of the models.

to Faster-RCNN ResNet-50 V1. However, the accuracy of Faster-RCNN ResNet-50 V1 is 12% higher than that of the Faster-RCNN NASNet model. Therefore, Faster-RCNN ResNet-50 V1 seems to be an optimal solution, capable of processing data quickly with a relatively high level of accuracy.

4.2 Models testing

The capabilities of the selected neural networks are presented using the data of two patients with the referenced labeling, whose data were not used to train the models. Fig. 6 shows the images with the red-marked area of stenosis. The models with the best values of the loss function and mAP were used for testing. Table 4 reports the best steps with the model optimal weights. The resultant predicting probability and location

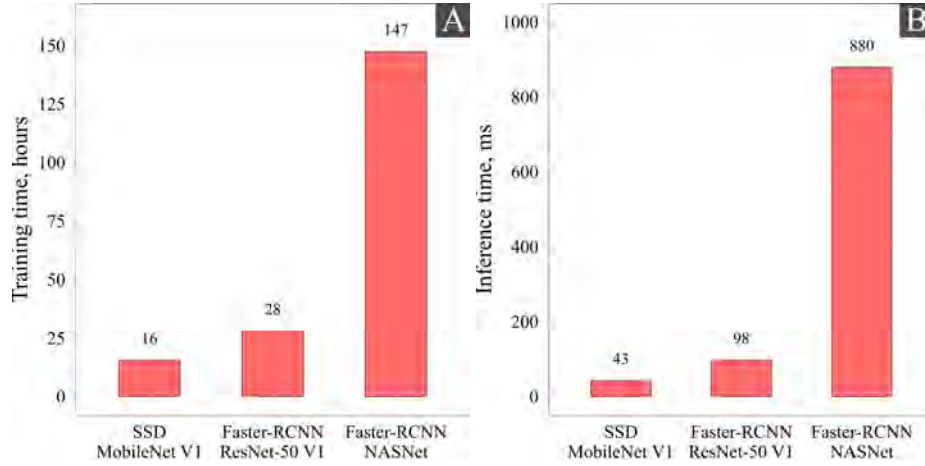


Fig. 5. Comparison of the training time (a) and inference time (b) of the models.

of stenosis are shown in Fig. 7. Additionally, Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics were used to assess the localization performance.

Table 4. Best steps with optimal model weights

Model	Best step
SSD MobileNet V1	24
Faster-RCNN ResNet-50 V1	84
Faster-RCNN NASNet	95

The comparative study proves that the models may accurately detect the location of stenosis. However, there were several false positives, while testing Faster-RCNN NASNet. In both cases, this model detected the location of the false stenotic segment in the right coronary artery and the anterior descending artery (two segments) besides the reference with a probability of over 90% (see Fig. 7c). However SSD MobileNet V1 erroneously predicted the absence of stenosis of the first patient. In addition, the efficiency of the detector based on the ResNet architecture, Faster-RCNN ResNet-50 V1, should be noted. The average DSC metric on the test data for this model was 0.85.

5 Conclusion

We trained three well-known and promising detectors based on different neural network architectures (MobileNet, ResNet-50, NASNet) to locate the single-vessel disease. Faster-RCNN ResNet-50 V1 has an optimal accuracy-to-speed ratio. This model

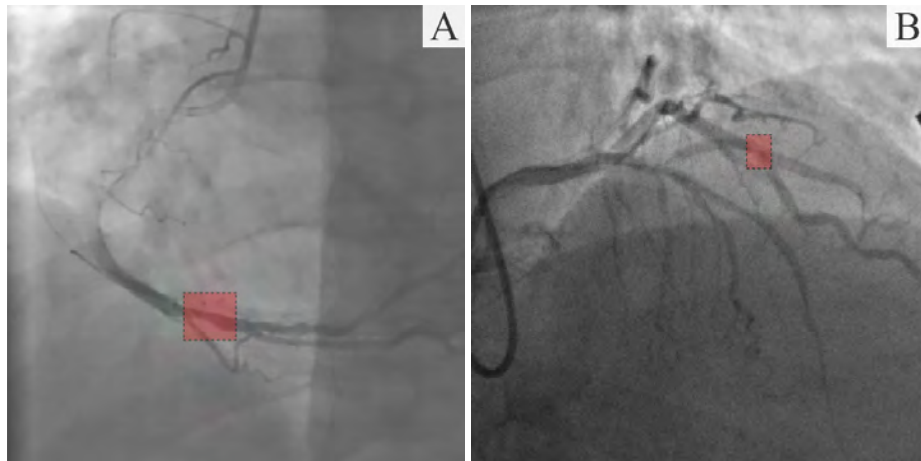


Fig. 6. Example of the reference test data.

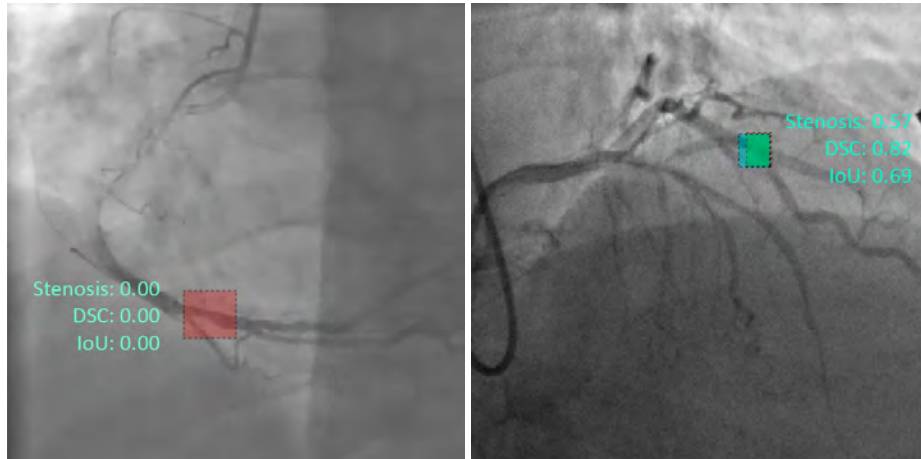
demonstrates the mean Average Precision (mAP@0.5 metric) of 0.92 at 10 frames per second. The fastest and relatively lightweight SSD MobileNet V1 model has the mean Average Precision of 0.70 at 23 frames per second. The ML-based approach proposed in this study is of particular interest in detecting the location of multivessel coronary artery disease. This approach ensures accurate detection of the stenosis location and may provide additional characteristics of the stenotic segment, such as its length, diameter, lateral branches, etc.

6 Acknowledgements

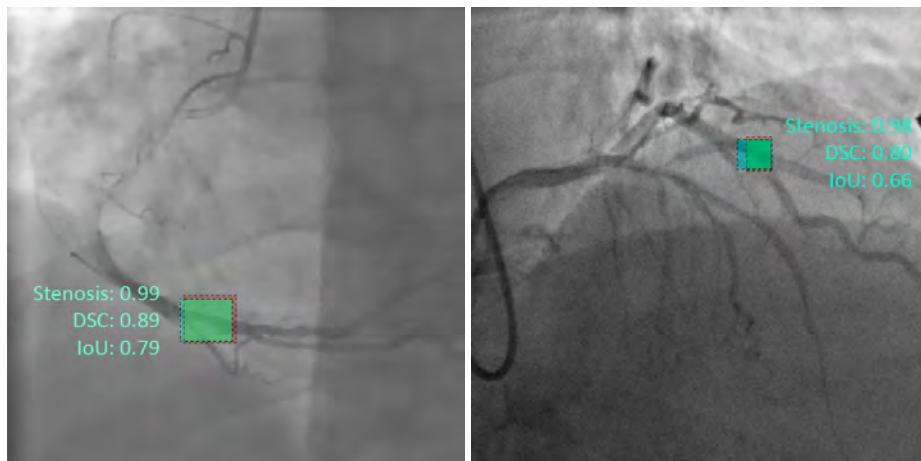
Data mining, data pre-processing and development of the ML-based approach to detect stenosis were supported by a grant from the Russian Science Foundation, project No. 18-75-10061 "Research and implementation of the concept of robotic minimally invasive prosthetics of the aortic valve". The training of the developed models using Amazon Web Services was funded by the Ministry of Science and Higher Education, project No. FFSWW-2020-0014 "Development of the technology for robotic multiparametric tomography based on big data processing and machine learning methods for studying promising composite materials". The selection of the primary metrics suggesting the model performance and their analysis was supported by the grant of the Russian Foundation of Basic Research, project No. 19-07-00351/19 "Methods and intelligent technologies for the scientific justification of strategic solutions on digital transformation".

References

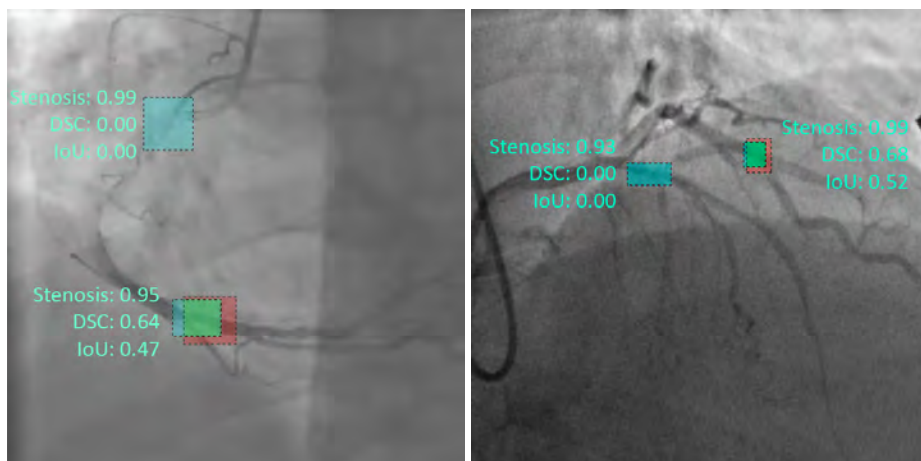
1. GBD 2017 Causes of Death Collaborators: Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a system-



(a) SSD MobileNet V1



(b) Faster-RCNN ResNet-50 V1



(c) Faster-RCNN NASNet

Fig. 7. Example of new data prediction.

- atic analysis for the Global Burden of Disease Study 2017. *Lancet* (London, England) **392**(10159), 1736–1788 (nov 2018). [https://doi.org/10.1016/S0140-6736\(18\)32203-7](https://doi.org/10.1016/S0140-6736(18)32203-7)
2. Antczak, K., Liberadzki, Ł.: Stenosis Detection with Deep Convolutional Neural Networks. *MATEC Web of Conferences* **210**, 04001 (oct 2018). <https://doi.org/10.1051/mateconf/201821004001>
 3. Chi, Y., Huang, W., Zhou, J., Toe, K.K., Zhang, J.M., Wong, P., Lim, S., Tan, R.S., Zhong, L.: Stenosis detection and quantification on cardiac CTCA using panoramic MIP of coronary arteries. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 4191–4194. IEEE (jul 2017). <https://doi.org/10.1109/EMBC.2017.8037780>
 4. Kang, D., Dey, D., Slomka, P.J., Arsanjani, R., Nakazato, R., Ko, H., Berman, D.S., Li, D., Kuo, C.C.J.: Structured learning algorithm for detection of nonobstructive and obstructive coronary plaque lesions from computed tomography angiography. *Journal of Medical Imaging* **2**(1), 014003 (mar 2015). <https://doi.org/10.1117/1.jmi.2.1.014003>
 5. Kang, D., Slomka, P.J., Nakazato, R., Arsanjani, R., Cheng, V.Y., Min, J.K., Li, D., Berman, D.S., Jay Kuo, C.C., Dey, D.: Automated knowledge-based detection of nonobstructive and obstructive arterial lesions from coronary CT angiography. *Medical Physics* **40**(4), 041912 (apr 2013). <https://doi.org/10.1118/1.4794480>
 6. Toledano, M., Lindenbaum, M., Lessick, J., Dragu, R., Ghersin, E., Engel, A.: Learning to Detect Coronary Artery Stenosis from Multi-Detector CT imaging. Tech. rep., Technion - Israel Institute of Technology, Haifa (2010)
 7. de Vos, B.D., Wolterink, J.M., Leiner, T., de Jong, P.A., Lessmann, N., Išgum, I.: Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. *IEEE transactions on medical imaging* **38**(9), 2127–2138 (sep 2019). <https://doi.org/10.1109/TMI.2019.2899534>
 8. Zreik, M., Van Hamersvelt, R.W., Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography. *IEEE Transactions on Medical Imaging* **38**(7), 1588–1598 (jul 2019). <https://doi.org/10.1109/TMI.2018.2883807>
 9. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single Shot MultiBox Detector. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905 LNCS, pp. 21–37. Springer Verlag (dec 2016). https://doi.org/10.1007/978-3-319-46448-0_2
 10. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(6), 1137–1149 (jun 2017). <https://doi.org/10.1109/TPAMI.2016.2577031>
 11. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K.: Speed/accuracy trade-offs for modern convolutional object detectors. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. vol. 2017-January, pp. 3296–3305. Institute of Electrical and Electronics Engineers Inc. (nov 2017). <https://doi.org/10.1109/CVPR.2017.351>
 12. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (apr 2017)
 13. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4510–4520. IEEE (jun 2018). <https://doi.org/10.1109/CVPR.2018.00474>
 14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2016-Decem, pp. 770–778. IEEE Computer Society (dec 2016). <https://doi.org/10.1109/CVPR.2016.90>

15. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 9908 LNCS, pp. 630–645. Springer Verlag (2016). https://doi.org/10.1007/978-3-319-46493-0_38
16. Zoph, B., Le, Q.V.: Neural Architecture Search with Reinforcement Learning. 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings (nov 2016)
17. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning Transferable Architectures for Scalable Image Recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8697–8710. IEEE (jun 2018). <https://doi.org/10.1109/CVPR.2018.00907>
18. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1440–1448. IEEE (dec 2015). <https://doi.org/10.1109/ICCV.2015.169>
19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). vol. 42, pp. 2999–3007. IEEE (oct 2017). <https://doi.org/10.1109/ICCV.2017.324>