Transformer graph network for coronary plaque localization in CCTA

Mario Viti, Hugues Talbot, Nicolas Gogin

TRANSFORMER GRAPH NETWORK FOR CORONARY PLAQUE LOCALIZATION IN CCTA

Mario Viti [⋆†], *Hugues Talbot* [⋆], *Nicolas Gogin*[†]

[⋆]CentraleSupélec, Université Paris-Saclay, Inria, 9 rue Joliot-Curie, F-91190 Gif-sur-Yvette France
[†]GE Healthcare 283 Rue de la Minière, 78530 Buc France

ABSTRACT

Coronary CT angiography (CCTA) is the only non-invasive imaging technique that reliably depicts the anatomic extent of Coronary Artery Disease (CAD). While occlusion remains a highly predictive indicator of major cardiovascular events (MACE), there is growing evidence that the presence and characteristics of coronary atherosclerosis provide additional prognostic information. In CCTA calcified plaques display high-intensity Hounsfield Units (HU) representative features while more complex representations characterize high-risk soft plaques. As such, accurate identification and quantification is burdensome and time consuming because of the limited temporal, spatial and contrast resolutions of X-ray scanners. Despite the success of deep learning in medical imaging, automatic localization of coronary plaques and especially soft plaques remains a challenging subject in CCTA vessel analysis. For this study, 150 CCTA scans were retrospectively collected. All patients were accepted at triage with minimal to severe CAD suspicion. Selection was carried out with uniform CAD-RADS severity distribution which normally follows an exponential decay function, thus obtaining a higher than normal concentration of plaques. The proposed method outperforms the state of the art for the localization of diverse types of plaques by exploiting the self-attention mechanism of transformers networks to embed contextual features of the coronary tree.

Index Terms— Medical image analysis, Computer-aided diagnosis, Angiorgraphic imaging, Heart, Vessels

1. INTRODUCTION

Coronary artery plaques present features of varying nature with corresponding functional values linked to CAD risks [1]. In recent studies, non calcified plaques were identified as a coronary obstruction independent factor associated with major adverse coronary events (MACE) [2]. A standard reference for automatic non invasive coronary analysis in CCTA is the FFR_{CT} [3][4] (Fractional Flow Reserve estimated via CT) which aims at simulating FFR invasive procedure used to predict vessel rupture risk and revascularization. This method, solely based on the segmentation of the coronary luminal volume, does not take into account anatomical plaque composition features. However in recent studies [5] it has been shown that a multivariate regression model including Calcium Score [6] and visually assessed anatomical plaque composition is as effective at predicting MACE as the FFR_{CT}score, while both have superior predictive power compared to the sole stenosis severity score. Advantageously, anatomical features have more reliable clinical interpretability, therefore plaque localization and characterization is necessary in order to achieve automatic clinical diagnosis.

1.1. Related methods

The 3D RCNN method [7] processes the curved planar reformat (CPR) of the straightened coronary volume [8] as a sequence of 3D cubes, each cube is processed individually by a 3DCNN and the sequence of extracted features is aggregated by a RNN. In this work the authors propose to characterize each manually annotated coronary segment with a label: whether it contains a plaque, its characteristic (calcified, mixed, soft) and stenosis (no-stenosis (0%), non-significant stenosis ($\geq 1\%$, $< 50\%$), occlusive stenosis ($\geq 50\%$)). To overcome the sequence classification shortcomings of poor localization the inference was performed on sliding fixed length sequences. In [9] the authors propose to avoid training the 3D CNN backbone in favor of Radiomics [10] feature extracted from the vessel wall segmentation, these features make use of the same RNN architecture for pooling information and were tested against an FFR functional score for validation. Subsequently in [11] the authors proposed to use as input two orthogonal views of the CPR volume to a 2D CNN classifier based on VGG-16 features extractor [12]. The method achieved similar results to the 3D RCNN in predicting revascularization and obstruction for a given input lesion, but with a more efficient pipeline.

2. METHOD

In our proposed method, $24 \times 24 \times 24$ 3D patches with isotropic resolution of 0.34 mm are sampled from the CCTA volume along the extracted centerline, each patch is processed individually by a 3D-CNN to extract local features.

The 3D-CNN is composed of 4 convolutional blocks interleaved with max pooling to achieve spatial reduction: $24^3 \mapsto 12^3 \mapsto 6^3 \mapsto 3^3$. Each convolutional block has 3 convolutional layers with residual connections and non-linear activation followed by batch normalization layers. The last convolutional layer is pooled by a global average pooling layer to obtain a fixed size feature vector (256 channels). The number of layers was tuned experimentally to be as small as possible without loss of performance.

We then exploit the directed tree structure in order to embed positional encodings in the features extracted. Given the graph $\mathcal{G} = (V, E), V = v_1, \ldots, v_n, E \in V \times V$ we begin with the definition of the *k*-hop (*k*-th order) neighborhood of the central node v_i as $\mathcal{N}_i = v_j \in V | d(i,j) \leq k$. If we were to consider an image as a 2D lattice grid a *k*-hop neighbourhood is what is usually referred to as a patch: We will define the layer operating on such neighbourhood as follows:

$$v_i^0 = v_i + pos_i$$

$$Q_i^l = \sigma(W_Q v_i^{l-1})$$

$$K_i^l = \sigma(W_K \mathcal{N}_i^{l-1})$$

$$V_i^l = \sigma(W_V \mathcal{N}_i^{l-1})$$

$$v_i^{l+1} = \text{Softmax}(Q_i^l {K_i^l}^\top / \sqrt{d_k}) V_i^l \quad (1)$$

Where W_K, W_Q, W_V are weight matrices multiplying each node feature vector $v_j \in \mathcal{N}_i$ and σ is a non linear activation function, where we overload the row vector multiplication to operate on sets as the result of $v^{l \geq 1}$ is the same independently of the ordering of $\mathcal{N}_i$. This operations are equivalent to the self-attention mechanism of the transformer architecture proposed in [13] which can efficiently model relations within a set of elements, we will also employ the technique of adding positional embedding pos_i as in eq (1). Positional encodings pos_i are computed locally by exploiting the directed tree structure using a simple signed hop count from the center node (+1 if distal, -1 if proximal). In [14] the authors propose using spectral embedding: however due to the arbitrary orientation of the eigenvector decomposition, the desirable property of a constant predecessor representation does not hold, so we choose to use the simple signed hop count. This is also preferable to a learnable positional encoder as it remains order-invariant.

The complete algorithmic pipeline is composed by the local features extractor 3D-CNN, these features are then added to the positional encodings to embed distal and proximal context. These features are merged within the Self Attention Block of the Spatial Transformer to obtain an attention vector and a prediction, the prediction being compared to the manually annotated label during training.

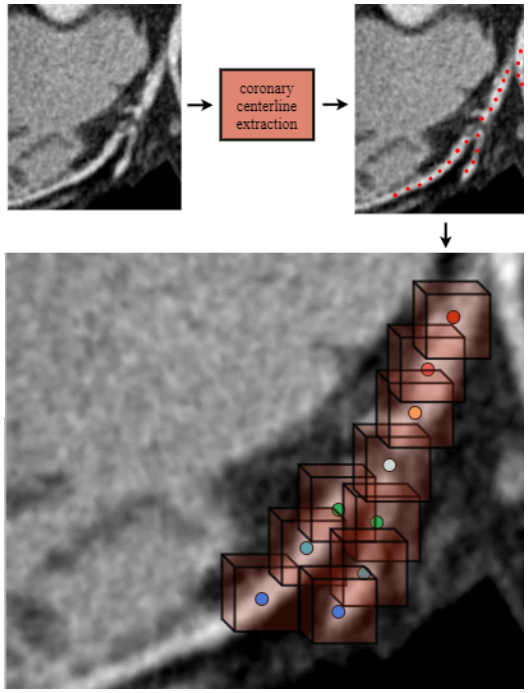

Fig. 1. CCTA image is preprocessed to extract the approximate locations along of the coronary centerline.

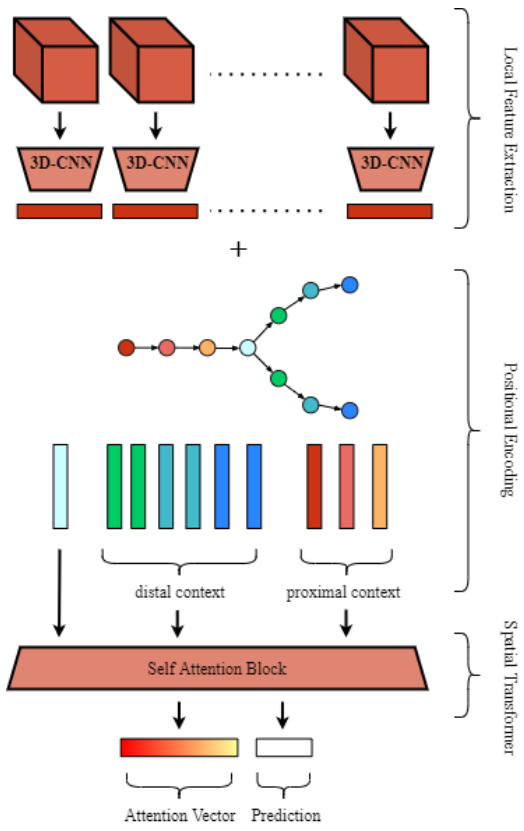

Fig. 2. Algorithmic pipeline for one location along the centerline.

3. EXPERIMENTS AND RESULTS

3.1. Training

Our model was trained in two separate sessions. First the 3D-CNN is trained to extract relevant features and is then frozen so that the Spatial Transformer can learn to merge the local image features and the positional encodings. The model is trained with a fixed batch size of 32 randomly sampled central nodes with corresponding neighbourhood for 200K steps using the Adam optimizer with automatic learning rate. This 2-stage strategy facilitates the notoriously difficult training of the Transformer [15].

3.2. Data

The dataset used in this work includes 150 coronary CT angiography (CCTA) scans, collected during a 2-year period from various clinical sites in France, USA and Italy. Images were acquired with tube voltage between 100 kVp and 120 kVp, a current between 600 mA and 1000 mA, a pixel spacing between 0.35 mm and 0.48 mm and 0.625 mm slice thickness. Each patient underwent both CCTA and anatomical tests [6] and was assigned a CAD-RADS [16] score by trained radiologists. The collection was carried out to obtain a CAD-RADS ≤ 4 uniform distribution: 24, 38, 30, 33, 25 patients with CAD-RADS from 0 to 4 respectively. This sample contains a higher than normal concentration of plaques as CAD-RADS distribution usually follows an exponentially decaying function. Each scan is paired with annotated coronary centerlines: for each lesion a starting and ending point is marked so that each point of the centerline is associated to a label: 0 (no-plaque), 1 (calcified plaque), 2 (mixed plaque), 3 (soft plaque) and a stenosis degree 0 (non-occlusive) 1 (occlusive). All annotations are based on clinical reports issued during medical review. Because of the impact of image quality [17] manual annotations were carried out to match the clinical report and the final annotations were submitted to a trained cardiologist for review. A subset of 50 patients with 224 annotated segments with plaques (49 soft, 78 mixed, 97 calcified plaques, and 182 non-occlusive, 42 occlusive stenosis) was chosen for testing using random stratified sampling.

3.3. Evaluation

The presented method is evaluated on the task of plaque localization. The classes were reduced to 0 (no-plaque) and 1 (calcified, mixed, soft plaque and occlusive and non-occlusive stenosis) or healthy and non-healthy or simply no-plaque and plaque. For the evaluation we choose to compare against [7] by excluding non suited methods present in literature. In [11] the method is provided with the starting and ending point to pad or stretch the plaque to a fixed window size, therefore it cannot be used for localization. In [9] the method bases the

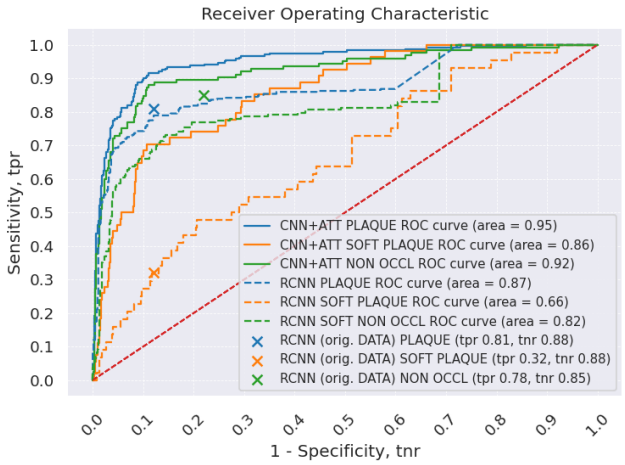

Fig. 3. ROC measurements. It must be noted that original data non occlusive stenosis comprises only $\geq 1\% < 50\%$ stenosis (non-significant) and not $< 50\%$ stenosis.

prediction on the radiomics features extracted from the lumen and vessel wall segmentation, it is therefore not suited for the proposed pipeline as we do not rely on the segmentation results. In [7] the method addresses plaque localization directly and solely relies on centerline extraction and the CPR straightened coronary volume. The original data test set from [7] comprises 65 patients with 191 annotated segments with plaques (28 Soft, 64 Mixed, 99 Calcified and which 155 non-occlusive, 36 occlusive stenosis). In order to compare with the state of the art performances is evaluated at the segment level. To avoid the bias of manually defined segments, we define segments as a contiguous coronary tract between 2 bifurcations, only segments with length ≤ 3.0 mm are excluded. We will adopt the same criteria used in [7] in which a true negative is a segment containing no positive labels, a true positive has labels which overlaps predictions. The maximum score for a given overlapping prediction over a segment is used to compute ROC measurements (fig. 3). It must be noted that in our dataset we define all non occlusive plaques with $< 50\%$ stenosis including 0% stenosis while the original data of [7] the non occlusive plaque was considered $\geq 1\% < 50\%$ stenosis (non-significant). We tested localization using 3 subsets of data: all data, soft plaque and no plaque, non occlusive plaque and no plaque, the results are reported in tab. 1. We conducted an ablation evaluation between CNN+ATT and CNN+GNN (Graph Neural Network) using weighted adjacency matrix (tab. 2).

4. DISCUSSION AND CONCLUSION

We presented a study focused on coronary plaque localization. In the literature coronary analysis methods tend to focus on different tasks related to coronary analysis (occlusion [9],

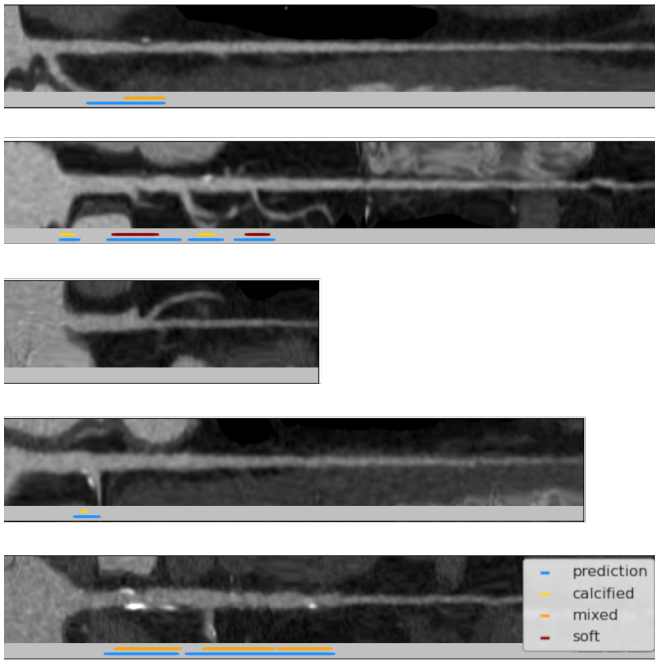

Fig. 4. A sample of the inference result displayed along the coronary with CPR reconstruction. For each picture the top and bottom bars represent the ground truth prediction (calcified, mixed and soft) overlapped with prediction of CNN+ATT.

Plaque	Data	Sens	Spec	Prec	F1	G-m
CNN+ATT	Ours	**0.89**	**0.90**	0.80	**0.84**	**0.89**
RCNN [7]	Ours	0.76	0.88	0.78	0.77	0.81
RCNN [7]	Orig	0.81	0.88	**0.81**	0.84	0.84
Soft Plaque	Data	Sens	Spec	Prec	F1	G-m
CNN+ATT	Ours	**0.70**	**0.90**	**0.4**	**0.50**	**0.78**
RCNN [7]	Ours	0.34	0.88	0.15	0.20	0.54
RCNN [7]	Orig	0.32	0.88	0.25	0.28	0.53
Non-occlus	Data	Sens	Spec	Prec	F1	G-m
CNN+ATT	Ours	**0.85**	**0.90**	0.65	0.74	**0.88**
RCNN [7]	Ours	0.71	0.88	0.60	0.65	0.79
RCNN [7]	Orig	0.78	0.85	**0.74**	**0.76**	0.82

Table 1. F1 $= 2(Prec * Sens)/(Prec + Sens)$ is an aggregated of Precision and Sensitivity , while G-m $= \sqrt{Spec * Sens}$ is an aggregated of Sensitivity and Specificity

AUC	PLAQUE	SOFT	NON-OCCL
CNN+ATT	**0.95**	**0.86**	**0.92**
CNN+GNN	0.94	0.82	0.90

Table 2. During ablation the CNN+ATT proposed method performed better than CNN+GNN

revascularization [11]) thus making comparing methodologies complex. For this work we decided to group all of these into a binary task: localizing healthy vs non-healthy (no-plaque, plaque) coronary segments. As previously discussed, non-occlusive ($< 50\%$ stenosis) and soft plaques are clinically relevant and therefore an automated diagnostic support tool must be able to localize them prior to characterization. The absence of coronary plaque sensibly reduces the risk of MACE, thus achieving reliable plaque localization could lead to a cost effective automatic patient triage, mitigating the burden on healthcare systems. The presented method outperforms the state of the art for CCTA soft, non-occlusive plaque localization. Unlike other methods for coronary analysis, the presented method does not rely on a straightened representation of coronaries and therefore can model bifurcations seamlessly. However [7] has good localization capability for a weakly supervised method making it less susceptible to incorrect annotations. [11] can predict revascularization quite efficiently (with 2D slices), when location of the lesion of interest is provided: therefore plaque characterization can be seen as a downstream task. These results show how soft plaque detection remains a highly challenging task. The model decision boundary between healthy and soft plaque tissue is fuzzy, especially in cases with absence of stenosis. This impacts on precision (tab. 1) because of false positive detection in the majority class, however the original dataset from [7] has a different class balance. Hard cases mining and the use of a focal loss can address this issue in an unsupervised manner and identify anomalous cases. We regard these as future directions to model and develop automated anatomical coronary analysis tools.

5. COMPLIANCE WITH ETHICAL STANDARD

5.1. Human rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

5.2. Informed consent and patient details

The authors declare that neither this report nor the collected data contain any personal information that could lead to the identification of the patient.

6. REFERENCES

[1] Udo Hoffmann, Maros Ferencik, and Udelson, "Prognostic Value of Noninvasive Cardiovascular Testing in Patients With Stable Chest Pain: Insights From the PROMISE Trial (Prospective Multicenter Imaging

Study for Evaluation of Chest Pain)," *Circulation*, vol. 135, no. 24, pp. 2320–2332, June 2017.

[2] Michelle C. Williams, Alastair J. Moss, and Marc Dweck, "Coronary Artery Plaque Characteristics Associated With Adverse Outcomes in the SCOT-HEART Study," *Journal of the American College of Cardiology*, vol. 73, no. 3, pp. 291–301, Jan. 2019.

[3] Ji Hyun Lee, Dalio Institute of Cardiovascular Imaging, New York-Presbyterian Hospital, New York, NY, USA, and Bríain ó Hartaigh, "Fractional Flow Reserve Measurement by Computed Tomography: An Alternative to the Stress Test," *Interventional Cardiology Review*, vol. 11, no. 2, pp. 105, 2016.

[4] Gianluca Pontone, Daniele Andreini, and Andrea I. Guaricci, "Rationale and design of the PERFECTION (comparison between stress cardiac computed tomography PERfusion versus Fractional flow rEserve measured by Computed Tomography angiography In the evaluation of suspected cOroNary artery disease) prospective study," *Journal of Cardiovascular Computed Tomography*, vol. 10, no. 4, pp. 330–334, July 2016.

[5] Wijnand J. Stuijfzand, Alexander R. van Rosendael, and Fay Y. Lin, "Stress Myocardial Perfusion Imaging vs Coronary Computed Tomographic Angiography for Diagnosis of Invasive Vessel-Specific Coronary Physiology: Predictive Modeling Results From the Computed Tomographic Evaluation of Atherosclerotic Determinants of Myocardial Ischemia (CREDENCE) Trial," *JAMA Cardiology*, vol. 5, no. 12, pp. 1338, Dec. 2020.

[6] Arthur S. Agatston, Warren R. Janowitz, and Hildner, "Quantification of coronary artery calcium using ultrafast computed tomography," *Journal of the American College of Cardiology*, vol. 15, no. 4, pp. 827–832, Mar. 1990.

[7] Majd Zreik, Robbert W. van Hamersvelt, and Jelmer M. Wolterink, "A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1588–1598, July 2019.

[8] Armin Kanitsar, Dominik Fleischmann, and Rainer Wegenkittl, "Diagnostic Relevant Visualization of Vascular Structures," in *Scientific Visualization: The Visual Extraction of Knowledge from Data*, Georges-Pierre Bonneau, Thomas Ertl, and Gregory M. Nielson, Eds., pp. 207–228. Springer-Verlag, Berlin/Heidelberg, 2006.

[9] Felix Denzinger, Michael Wels, Nishant Ravikumar, et al., "Coronary Artery Plaque Characterization from CCTA Scans Using Deep Learning and Radiomics," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Dinggang Shen, Tianming Liu, and Terry M. Peters, Eds., vol. 11767, pp. 593–601. Springer International Publishing, Cham, 2019.

[10] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, and Sara Carvalho, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[11] Felix Denzinger, Michael Wels, and Katharina Breininger, "Deep Learning Algorithms for Coronary Artery Plaque Characterisation from CCTA Scans," in *Bildverarbeitung für die Medizin 2020*, Thomas Tolxdorff, Thomas M. Deserno, Heinz Handels, Andreas Maier, Klaus H. Maier-Hein, and Christoph Palm, Eds., pp. 193–198. Springer Fachmedien Wiesbaden, Wiesbaden, 2020.

[12] M. Cimpoi, Subhransu Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3828–3836, 2015.

[13] Ashish Vaswani, Noam M. Shazeer, and Niki Parmar, "Attention is all you need," *ArXiv*, vol. abs/1706.03762, 2017.

[14] Vijay Prakash Dwivedi and Xavier Bresson, "A generalization of transformer networks to graphs," *CoRR*, vol. abs/2012.09699, 2020.

[15] Liyuan Liu, Xiaodong Liu, and Jianfeng Gao, "Understanding the Difficulty of Training Transformers," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, 2020, pp. 5747–5763, Association for Computational Linguistics.

[16] Harvey S. Hecht, Paul Cronin, and Michael J. Blaha, "2016 SCCT/STR guidelines for coronary artery calcium scoring of noncontrast noncardiac chest CT scans: A report of the Society of Cardiovascular Computed Tomography and Society of Thoracic Radiology," *Journal of Cardiovascular Computed Tomography*, vol. 11, no. 1, pp. 74–84, Jan. 2017.

[17] Farhood Saremi and Stephan Achenbach, "Coronary Plaque Characterization Using CT," *American Journal of Roentgenology*, vol. 204, no. 3, pp. W249–W260, Mar. 2015.