Automatic Coronary Calcium Scoring in Cardiac CT Angiography Using Convolutional Neural Networks

Jelmer M. Wolterink[1,⋆], Tim Leiner[2], Max A. Viergever[1], and Ivana Išgum[1]

[1] Image Sciences Institute, UMC Utrecht, Utrecht, The Netherlands
[2] Department of Radiology, UMC Utrecht, Utrecht, The Netherlands

Abstract. The amount of coronary artery calcification (CAC) is a strong and independent predictor of cardiovascular events. Non-contrast enhanced cardiac CT is considered a reference for quantification of CAC. Recently, it has been shown that CAC may be quantified in cardiac CT angiography (CCTA). We present a pattern recognition method that automatically identifies and quantifies CAC in CCTA. The study included CCTA scans of 50 patients equally distributed over five cardiovascular risk categories. CAC in CCTA was identified in two stages. In the first stage, potential CAC voxels were identified using a convolutional neural network (CNN). In the second stage, candidate CAC lesions were extracted based on the CNN output for analyzed voxels and thereafter described with a set of features and classified using a Random Forest. Ten-fold stratified cross-validation experiments were performed. CAC volume was quantified per patient and compared with manual reference annotations in the CCTA scan. Bland-Altman bias and limits of agreement between reference and automatic annotations were -15 (-198–168) after the first stage and -3 (-86 – 79) after the second stage. The results show that CAC can be automatically identified and quantified in CCTA using the proposed method. This might obviate the need for a dedicated non-contrast-enhanced CT scan for CAC scoring, which is regularly acquired prior to a CCTA scan, and thus reduce the CT radiation dose received by patients.

Keywords: Automatic coronary artery calcium scoring, Cardiac CTA, Convolutional neural network, Random Forest.

1 Introduction

Cardiovascular disease (CVD) is the global leading cause of death. The amount of coronary artery calcification (CAC) is a strong and independent predictor of CVD events, which can be identified and quantified in cardiac CT [1]. In clinical practice, CAC is routinely quantified using non-contrast enhanced, calcium scoring CT (CSCT) [2]. Recently, it has been shown that CAC may also be quantified in contrast-enhanced cardiac CT angiography (CCTA). Consequently,

⋆ This work has been financially supported by PIE Medical Imaging.

N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 589–596, 2015.
DOI: 10.1007/978-3-319-24553-9_72

a dedicated CSCT scan, which is often routinely acquired prior to CCTA, might potentially be omitted. This could reduce the radiation dose of a typical cardiac CT examination by 40-50% [3]. CAC in CSCT can be identified manually by an expert or automatically [4,5]. In both situations, a threshold of 130 HU is used to identify connected voxels representing CAC. This method is not generalizable to CCTA. The coronary artery lumen is typically enhanced beyond 130 HU, which makes differentiation of CAC and lumen challenging. Other global attenuation thresholds for manual CAC scoring in CCTA have therefore been proposed [6,7]. However, a global threshold might limit the applicability of the method to scans acquired with different protocols, scanners or contrast agents. Alternatively, patient-specific attenuation thresholds were proposed, based on HU values taken from an ROI in the ascending aorta [8] or the proximal coronary arteries [9]. A drawback of these thresholds is the limited repeatability of user-defined ROIs. Furthermore, the large number of image slices in CCTA (> 200) makes routine manual scoring of CAC in CCTA time-consuming.

To overcome these limitations, automatic methods have been proposed. Previously published automatic methods were based on a (semi)-automatically extracted segmentation of the coronary arteries. This segmentation was used to identify CAC as deviations from a trend line through the lumen intensity [10,11], deviations from a model of non-calcified artery segments [12], or as voxels in the extracted arteries with intensities above a patient-specific HU threshold [13]. These methods have shown good performance, but depend on successful coronary artery extraction. This might fail in patients with complex anatomy, in the distal segments of the coronary arteries or in scans with motion or noise artifacts. In addition, severe CAC deposits affect the performance of artery extraction algorithms, restricting their applicability in CAC identification [14]. Failure of the extraction step would result in failure of the automatic CAC scoring method.

Therefore, we propose a novel pattern recognition based method, which identifies CAC in CCTA without artery extraction. The method uses a convolutional neural network (CNN) for the identification of potential CAC voxels and subsequently a Random Forest for the identification of CAC lesions. Previous work has demonstrated that CNNs can provide insights via automatically derived feature hierarchies leading to highly accurate results [15,16]. In this paper, we combine texture features that the CNN automatically derives from image patches with basic location features that are extracted from the input image. Automatically derived CAC scores are compared with manual annotations in CCTA and CSCT.

2 Material and Methods

2.1 Data

We retrospectively inspected 116 CT examinations of consecutively scanned patients for whom both a CCTA and a CSCT scan were available. CAC scores (Agatston scores) had previously been manually determined in the CSCT scans by experts. Based on these scores, ten consecutively scanned patients from each

of five CVD risk categories (I:0, II:1-10, III: 11-100, IV:101-400, V:>400) [1] were included in the data set. The CCTA of one additional patient with CAC score zero was included as an atlas. All CCTA scans were acquired on a 256-detector row scanner (Philips Brilliance iCT, Philips Medical, Best, The Netherlands) using 120 kVp and 210-300 mAs, with ECG-triggering and contrast enhancement. Reconstructed sections had 0.45 mm spacing, 0.90 mm thickness and 0.4-0.5 × 0.4-0.5 mm in-plane resolution.

To train and evaluate the method, manual reference annotations in the CCTA scans were obtained as follows. A region of interest in the ascending aorta was manually defined. The mean HU_{aorta} and standard deviation SD of the intensities in this region were used to compute a patient-specific threshold HU_{aorta} + 2 SD [8]. This threshold was locally adjusted to correct for erroneous inclusion of contrast-enhanced lumen. CAC lesions were manually annotated with region growing in the thresholded image.

2.2 Automatic CAC Scoring

The proposed method for automatic CAC scoring consisted of two stages: a voxel classification stage and a lesion classification stage. In the voxel classification stage, potential CAC locations were identified using a CNN. Based on the output of the CNN, 3D-connected voxels with a high probability of being CAC (p_{CAC}) were extracted as lesions. These lesions were described by a set of features and classified as CAC or non-CAC using a Random Forest classifier. Identified CAC volume and Agatston score were quantified per patient.

Candidates in the voxel classification stage were voxels with intensity above a patient-specific threshold θ_{PS}, the 98th percentile of the CCTA HU intensity histogram. This is a conservative threshold which correlated well (Spearman's ρ 0.89) with that proposed in [8]. Determination of this threshold required no user interaction. Each voxel was represented by three 24 × 24 patches from orthogonal planes centered at the voxel [15,16]. The patch size was chosen to be large enough to contain CAC lesions of moderate size or the surrounding coronary artery lumen, while remaining within the computational limitations of our hardware. Figure 1 shows an example candidate with orthogonal patches (bottom left). Because patches were small, they did not convey much anatomical information. Therefore, images were registered to the atlas CCTA using affine and elastic registration with *elastiX* [17] and parameters as in [5], and voxel x, y and z-coordinates in the atlas image were computed as location features.

A CNN with two convolutional and max-pooling layers, two fully connected hidden layers and a softmax output layer was used (Figure 1). Features were derived from the orthogonal input patches in the convolutional layers and combined with the three location features as input to the first hidden layer. All units used the rectifier activation function [18]. The network was trained with mini-batch learning and RMSprop, which uses a moving average of recent gradients to normalize the gradient in each iteration [19]. We sought to obtain translation invariant features by max-pooling after each convolutional layer. To prevent overfitting, Dropout was used to simulate the training of a large number

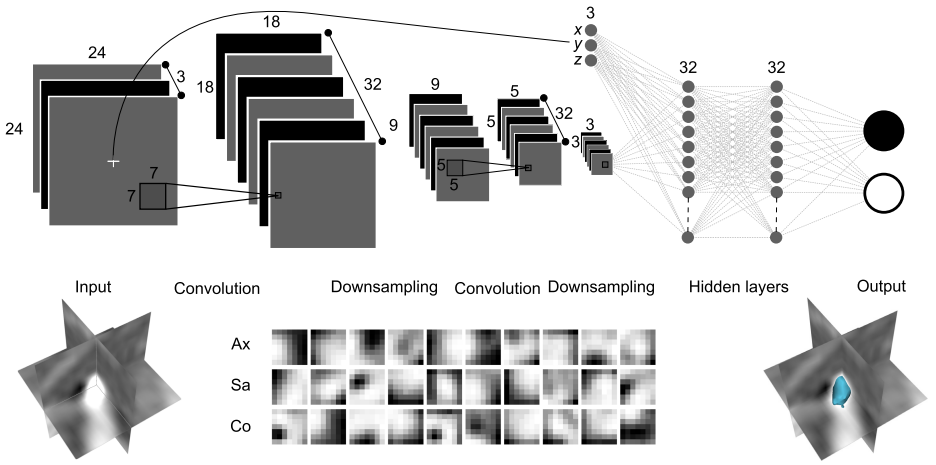


Fig. 1. *Top* CNN architecture. The CNN had two convolutional layers with max-pooling, two fully connected hidden layers and one softmax output layer. Location features (x, y, z) were additional input to the first hidden layer. *Bottom* Example input CAC voxel representation: three orthogonal input patches of size 24×24. Convolutional filters for axial (Ax), sagittal (Sa) and coronal (Co) input patches. Rendering of thresholded output for candidate and neighboring voxels.

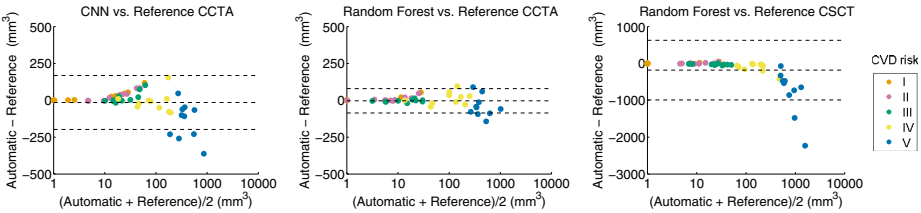
of thinned networks [20]. The softmax output layer of the CNN returned p_{CAC}: the probability of a candidate voxel to belong to a CAC lesion.

In the lesion classification stage, p_{CAC} was first thresholded to consider only voxels that were likely CAC. The threshold value ($p_{CAC} \geq 0.75$) was set based on precision-recall analysis of the voxel classification, so that most CAC voxels were retained and individual lesions could be extracted by 3D-connected component labeling. Each extracted lesion was described by 14 features. The following statistics of the p_{CAC} values of the voxels in the candidate lesion were features: minimum, maximum, range, average and standard deviation. These features were supplemented with features that have been shown to benefit CAC detection in CSCT [4,5]: lesion volume and the minimum, maximum, range, average and standard deviation of intensity values of voxels in the lesion. Furthermore, x, y and z-coordinates of lesions in the atlas image were used as location features. All features were used as input to a Random Forest classifier [21].

3 Experiments and Results

For both CNN and Random Forest classification, experiments were performed using stratified ten-fold cross validation. Each fold contained one patient in each of five CVD risk categories.

To train and test the CNNs, approximately $800,000$ candidate voxels were extracted per patient. From these, 15,000 candidates were selected for training,



(a) (b) (c)

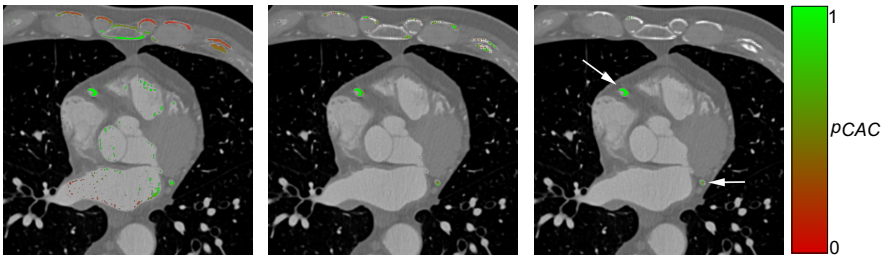
Fig. 2. Bland-Altman plots comparing reference and automatic per patient CAC volume in CCTA. Automatic results after (a) CNN output thresholding and (b) Random Forest lesion classification. In (c), Random Forest lesion classification is compared with reference CAC volume in CSCT. Patients are color-coded according to their CSCT-based CVD risk category.

including all positives (CAC) and a random selection of negatives. Hence, for each fold a CNN was trained with $45 \times 15,000 = 675,000$ candidates using mini-batch training with batches of 512 samples. The Dropout retention probability was 0.5 for all units. CNNs were implemented in Theano [22] and run on single Nvidia GRID K520 GPUs. For each fold, a separate CNN was trained for 30 epochs with a training time of 4,5 hours. Voxel classification for a single scan took approximately ten minutes. Figure 1 shows convolution filters trained in the first layer of one of the trained CNNs.

To train and test the Random Forest a total of 2268 candidate lesions were extracted by thresholding at $p_{CAC} \geq 0.75$ and 3D-connected component labeling of the thresholded image. Among candidates, 253 were positives, constituting 30% of candidate volume. All extracted candidates were used for training. Hence, for each fold a Random Forest was trained with approximately 2041 candidates. The four most important features in the Random Forest classification were the mean, maximum and standard deviation of voxel intensities and the minimum CNN output value p_{CAC} in a candidate.

The output of the CNN (p_{CAC}) may be directly thresholded for voxel classification and CAC quantification. Thresholding at $p_{CAC} \geq 0.95$ resulted in Bland-Altman mean and limits of agreement of -15 (-198–168) (Figure 2a). Lesion extraction and Random Forest classification improved these values to -2 (-86 –79) (Figure 2b). However, CAC volume was overestimated for several patients in the lower two CVD risk categories, including one patient with a calcified mediastinal lymph node and one patient with a CAC-like lesion near the sternum.

Figure 2c shows a comparison of automatically determined CAC volumes in CCTA and reference CAC volumes in CSCT. CAC volume and Agatston scores were substantially lower in CCTA than in CSCT, which agrees with previously published studies on manual CAC scoring in CCTA [8,23]. Nevertheless, CAC quantification in CCTA can differentiate between patients with zero



(a) (b) (c)

Fig. 3. Output of three CNNs on example image slice. (a) Using only location information, candidates near the center of the scan were favored. (b) Using only patch values, ribs were erroneously identified as CAC. (c) Using both patch values and location information, only true CAC in the right and left circumflex coronary artery (arrows) was detected.

CAC and non-zero CAC. Namely, the method correctly identified 8/10 patients with zero CAC according to CSCT and 36/40 patients with non-zero CAC as determined with CSCT. In addition, the method correctly identified 9/10 patients at very high risk of a CVD event (Agatston score > 400) according to CSCT and 40/40 patients at low or intermediate risk of a CVD event (Agatston score ≤ 400).

The CNNs used features derived from the input patches, as well as information about the location of candidates. To evaluate the effect of these location features on CAC identification, we trained two additional CNNs for one fold. One CNN was trained with patch values set to zero, but location features preserved. A second CNN was trained with location features set to zero, but patch values preserved. Figure 3 shows that the CNN that only used location features (Figure 3a) favored candidates in the heart, but made no distinction between true CAC and attenuated blood. The CNN that only used patch input (Figure 3b) identified calcific objects, but made no distinction between CAC and ribs. The full CNN (Figure 3c) used both location and texture information to correctly identify two CAC lesions in the right and left circumflex coronary arteries.

4 Discussion and Conclusion

A pattern recognition method for the automatic identification of CAC in CCTA has been described. The method uses CNNs to locate likely CAC voxels and Random Forests to classify extracted CAC lesions. To our knowledge, this is the first application of CNNs to CAC scoring in CT. In contrast to other automatic CAC scoring methods, it does not rely on automatic coronary artery extraction.

The results showed that CNNs were able to identify voxels likely to be CAC. Combining automatically derived patch-based texture features and candidate location features proved advantageous. However, the use of a single atlas for location determination might introduce bias towards a single image. In future

work, we will investigate a multi-atlas based coordinate space, which could also aid candidate voxel extraction. Extraction was now based on HU thresholds, similar but not equal to those for manual reference scoring. Hence, this could cause discrepancies between reference and automatic volumes. Furthermore, in the current study, candidate voxels were represented by three orthogonal patches instead of a full 3D representation. A comparable full 3D representation with 24 voxels along each dimension might contain more information, but also increases the input size eight-fold. The size of the model increases accordingly, posing memory requirements which we found could not be met by our hardware. In addition, the larger number of weights might require more training samples than provided in the current study. It has previously been shown that smaller 3D representations can be outperformed by larger sparse orthogonal patches [15]. The voxel classification output of the CNNs was enhanced by Random Forest classification of extracted lesions. This resulted in good agreement between reference and automatic per patient CAC volumes.

In conclusion, CAC can be automatically identified and quantified in CCTA using the proposed pattern recognition method. This might obviate the need for a dedicated CSCT for CAC scoring, which is regularly acquired prior to a CCTA, and thus reduce the CT radiation dose received by patients.

References

1. Rumberger, J.A., Brundage, B.H., Rader, D.J., Kondos, G.: Electron beam computed tomographic coronary calcium scanning: a review and guidelines for use in asymptomatic persons. Mayo Clin. Proc. 74(3), 243–252 (1999)
2. Hecht, H.S.: Coronary artery calcium scanning: Past, present, and future. JACC: Cardiovasc. Imag. 8(5), 579–596 (2015)
3. Voros, S., Qian, Z.: Agatston score tried and true: by contrast, can we quantify calcium on CTA? J. Cardiovasc. Comput. Tomogr. 6(1), 45–47 (2012)
4. Shahzad, R., van Walsum, T., Schaap, M., Rossi, A., Klein, S., Weustink, A.C., de Feyter, P.J., van Vliet, L.J., Niessen, W.J.: Vessel specific coronary artery calcium scoring: An automatic system. Acad. Radiol. 20(1), 1–9 (2013)
5. Wolterink, J.M., Leiner, T., Takx, R.A.P., Viergever, M.A., Išgum, I.: Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. IEEE Trans. Med. Imag. 34(9), 1867–1878
6. Otton, J.M., Lønborg, J.T., Boshell, D., Feneley, M., Hayen, A., Sammel, N., Sesel, K., Bester, L., McCrohon, J.: A method for coronary artery calcium scoring using contrast-enhanced computed tomography. J. Cardiovasc. Comput. Tomogr. 6(1), 37–44 (2012)
7. Glodny, B., Helmel, B., Trieb, T., Schenk, C., Taferner, B., Unterholzner, V., Strasak, A., Petersen, J.: A method for calcium quantification by means of CT coronary angiography using 64-multidetector CT: very high correlation with Agatston and volume scores. Eur. Radiol. 19(7), 1661–1668 (2009)
8. Mylonas, I., Alam, M., Amily, N., Small, G., Chen, L., Yam, Y., Hibbert, B., Chow, B.J.: Quantifying coronary artery calcification from a contrast-enhanced cardiac computed tomography angiography study. Eur. Heart J. Cardiovasc. Imaging 15(2), 210–215 (2014)

9. Pavitt, C.W., Harron, K., Lindsay, A.C., Ray, R., Zielke, S., Gordon, D., Rubens, M.B., Padley, S.P., Nicol, E.D.: Deriving coronary artery calcium scores from CT coronary angiography: a proposed algorithm for evaluating stable chest pain. Int. J. Card. Imaging 30(6), 1135–1143 (2014)
10. Wesarg, S., Khan, M.F., Firle, E.A.: Localizing calcifications in cardiac CT data sets using a new vessel segmentation approach. J. Digit. Imaging 19(3), 249–257 (2006)
11. Ahmed, W., de Graaf, M.A., Broersen, A., Kitslaar, P.H., Oost, E., Dijkstra, J., Bax, J.J., Reiber, J.H., Scholte, A.J.: Automatic detection and quantification of the Agatston coronary artery calcium score on contrast computed tomography angiography. Int. J. Cardiovasc. Imaging 31(1), 151–161 (2014)
12. Eilot, D., Goldenberg, R.: Fully automatic model-based calcium segmentation and scoring in coronary CT angiography. IJCARS 9(4), 595–608 (2014)
13. Teßmann, M., Vega-Higuera, F., Bischoff, B., Hausleiter, J., Greiner, G.: Automatic detection and quantification of coronary calcium on 3D CT angiography data. CSRDC 26(1), 117–124 (2011)
14. Schaap, M., Metz, C.T., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N.R., Bauer, C., Bogunović, H., Castro, C., Deng, X., et al.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Medical Image Analysis 13(5), 701–714 (2009)
15. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013)
16. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 520–527. Springer, Heidelberg (2014)
17. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imag. 29(1), 196–205 (2010)
18. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, vol. 15, pp. 315–323 (2011)
19. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning 4 (2012)
20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15(1), 1929–1958 (2014)
21. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
22. Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I.J., Bergeron, A., Bouchard, N., Bengio, Y.: Theano: new features and speed improvements. In: Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop (2012)
23. van der Bijl, N., Joemai, R.M., Geleijns, J., Bax, J.J., Schuijf, J.D., de Roos, A., Kroft, L.J.: Assessment of Agatston coronary artery calcium score using contrast-enhanced CT coronary angiography. Am. J. Roentgenol. 195(6), 1299–1305 (2010)