Maximilian Eibl, Martin Gaedke (Hrsg.): INFORMATIK 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018 1

Automatic Aortic Wall Segmentation and Plaque Detection using Deep Convolutional Neural Networks

Marcel Beetz[1]

Abstract: Abnormal aortic wall thickness and the presence of aortic plaque have been linked to various types of cardiovascular disease. Quantification of both indicators currently depends on manual or semi-automatic methods which suffer from limited quality and long acquisition times. This work presents various fully automatic state-of-the art solutions to two medical image processing problems: aortic wall segmentation and plaque slice detection. A u-net derived residual convolutional neural network (CNN), a cascaded pipeline of two CNNs and a 3D CNN architecture are used for aortic wall segmentation. Plaque detection is performed by a standard multilayer residual CNN classification architecture, a u-net derived CNN classifier and a capsule CNN. The experiments show that the u-net inspired residual CNN performs best at the aortic wall segmentation task with a Dice score of around 0.8 while the capsule CNN achieves the best results in slice-wise plaque detection with a precision of 0.74 and an accuracy of 0.68.

Keywords: Deep Learning; Convolutional Neural Networks; Capsule Networks; Aortic Wall Segmentation; Plaque Detection; Medical Image Processing

1 Introduction

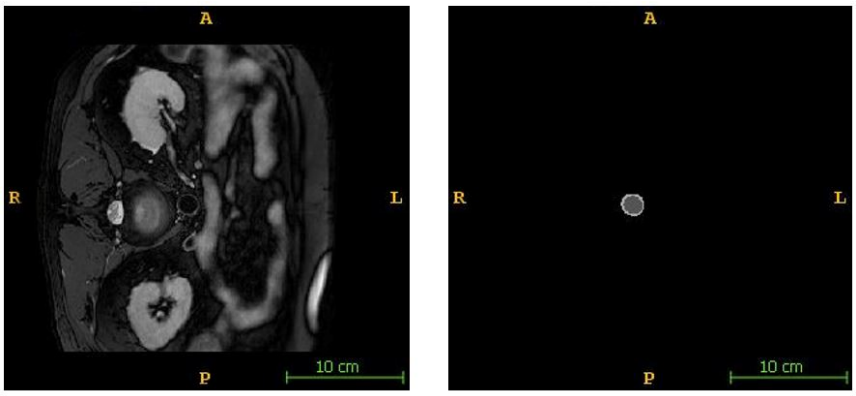
Cardiovascular disease continues to be the most common cause of death in Europe [Wi17]. Both aortic wall thickness and aortic plaque have been shown to be potential indicators in predicting and identifying cardiovascular disease [Gu10] [Ho16]. More specifically, abnormal aortic wall thickness has been linked to medical conditions such as atherosclerosis [Li15] while abdominal aortic plaque has been associated with coronary artery disease [Li16]. In both cases, there is limited automation in current medical practice with only semi-automatic methods being used to obtain segmentations or detect plaque. These semi-automatic techniques still require considerable human intervention leading to higher costs, human errors, inter- and intra-operator variability and lower reproducibility. In addition, medical images are generally more challenging to segment due to a smaller data set size, low and anisotropic image resolution, intensity inhomogeneities and the presence of imaging artifacts. This results in poor accuracy of traditional segmentation methods whose reliance on intensity or shape information alone is not able to capture the complexities of medical images, including those of the aorta region. Therefore, this paper proposes various fully

[1] Department of Informatics, Technical University of Munich, Boltzmannstr. 3 , 85748 Garching, Germany
marcel.beetz@tum.de

automatic methods derived from latest advances in deep learning to solve many of the aforementioned issues.
In summary the contributions of this paper are twofold:

- Three state-of-the-art fully automatic medical image segmentation methods, namely a u-net derived residual convolutional neural network (CNN), a cascaded CNN architecture and a 3D CNN, are applied to aortic wall segmentation.
- A capsule network as a recent innovation in general image classification is adapted for the detection of plaque in medical images and compared to previous state-of-the-art deep learning-based classifiers.

Both methods were applied and evaluated on a dataset of T2 abdominal images. Fig. 1 shows an example of a 512x512 input image of this dataset and the corresponding segmentation mask. The black area in the segmentation mask identifies the *background* region, while gray and white colored areas refer to the aorta's *blood-pool* and the *aortic wall* respectively. The aorta makes up only a small portion of the overall image but is still relatively easy to detect due to its recognizable shape. The aortic wall however only consists of a few pixels. This combined with the other previously mentioned challenges of medical data sets makes it difficult to find an accurate segmentation automatically. In plaque-containing slices, the *plaque* region is located on the inner side of the aortic wall with an average diameter of a few voxels and a partially circumferential shape.



a) Image b) Segmentation mask

Fig. 1: Example of a 512x512 input image (a) and its corresponding segmentation mask (b)

2 Previous work

Recently, deep convolutional neural networks have been shown to outperform traditional methods on a variety of data sets in both image segmentation and classification [Po11].

Common benchmarks such as MNIST for handwritten digit classification or the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [Ru15] to detect real-world objects from 200 categories have seen tremendous improvements achieved by deep learning-based methods. AlexNet [KSH12] was the first CNN to be applied to image classification, winning the 2012 ILSVRC by an error rate of more than 10% lower than the second best method. It introduced dropout layers and data augmentation to combat model overfitting. VGG net [SZ14] decreased the kernel size of the convolutional layers and showed the feasibility of very deep architectures. ResNet [He16] was the first model to overcome average human level accuracy with a classification error rate at ILSVRC of 3.6%. It introduced residual blocks to enable easier gradient flow and therefore network training allowing it to further increase its architectural depth. A recent advancement have been capsule CNNs which showed an equal and in some cases a better performance than current state-of-the-art methods on the MNIST dataset [SFH17]. In image segmentation, fully convolutional neural networks (FCN) [LSD15] enabled a pixel-to-pixel mapping from input image to the output segmentation mask thereby improving upon previous region-based segmentation methods. Specifically regarding medical images, the u-net [RFB15] allowed for CNNs to be used in a complete image segmentation pipeline by combining a down-sampling path with normal convolutional layers and an up-sampling path with transpose convolutions to obtain a segmentation mask with the same dimensions as the input image in one pass over the network. Further advances introduced short residual connections and more complex combinations of convolutional, normalization and activation layers [Dr16]. More recently, both cascaded segmentation pipelines [Ch17] and 3D convolutional neural networks [MNA16] have been used to improve medical image segmentation results even further. Deep CNNs have also been successfully applied in medical imaging as a classifier for various types of cancer (e.g. breast cancer [Ar17]). However, both aortic wall segmentation and plaque detection have seen limited use of deep learning-based methods and have been most commonly performed via semi-automatic approaches based on graph cuts [Du12] or geodesic active contour models [Wa17].

3 Methods

This section describes the general data preparation steps, the three approaches to each of the two main problems examined in this paper and the evaluation metrics.

3.1 Data preparation

Various widely used preprocessing steps were applied to the dataset to ease comparability of data instances among each other and improve the network's learning and prediction capabilities. First, voxel spacing was adjusted to a common value among all data instances and the few images with a different resolution than 512x512 were resampled to that value. Next, slice-wise normalization of intensity values was applied to facilitate the network's

learning process. In addition, histogram equalization was used in order to enhance image contrast. The segmentation masks associated with the raw images of wrong resolution were also resampled to 512x512 using the nearest neighbor method to maintain the whole numbers that identify each image region. Voxel spacing was adjusted for the segmentation masks as well. Data augmentation resulted in worsened performance compared to solely relying on the input data set and was therefore not used for the methods presented in the following.

3.2 Basic CNN architecture

The basic CNN architecture in this paper is derived from the popular u-net [RFB15] and work by Drozdzal et al. [Dr16] (Fig. 2 (a)). It consists of a downward contracting path to extract relevant features followed by an upward expanding path to recreate a predicted segmentation mask with the same dimensionality as the original input image. Each additional convolutional layer along the contracting path has the task of learning increasingly complex features. Long skip connections between the downward and upward path concatenate the respective feature maps to allow for better information retention between the paths. The dimensionality of the feature maps after each level is provided in Fig. 2 (a) with the first two numbers showing the feature map's x and y dimension and the third number depicting the number of feature maps used.

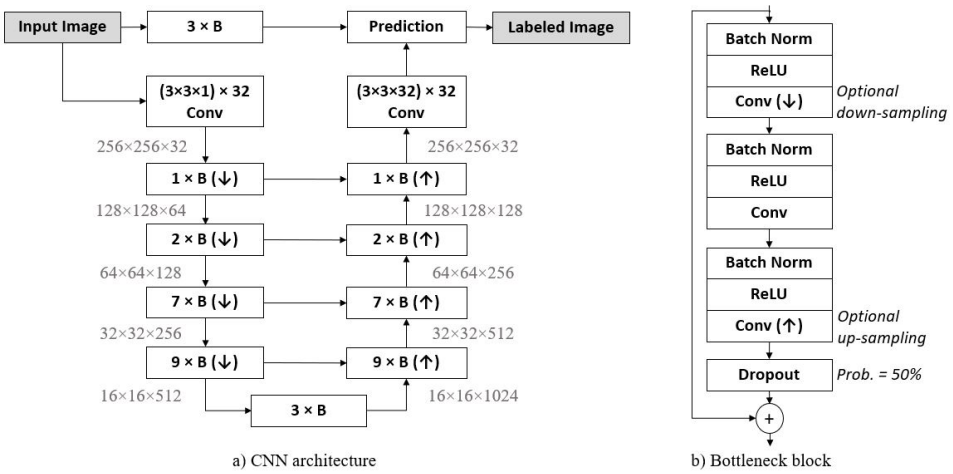


Fig. 2: Overview of complete CNN (a) and bottleneck block architecture (b)

Each layer in the CNN architecture consists of a number of bottleneck blocks which in return are made up of three iterations of a batch normalization, an activation and a convolutional layer followed by a dropout layer at the end (Fig. 2 (b)). Special convolutional layers with a stride value of 2 are used for down- or up-sampling the feature map resolution, which is indicated in Fig. 2 (b) by an up- or downward pointing arrow respectively. In addition, a

short residual connection is applied around the entire block to also allow information flow at a smaller scale (Fig. 2 (b)). Due to memory limitations, a 256x256 sized fixed ROI around the aorta is cropped out of the original image slices and used as an input to the CNN.

3.3 Cascaded CNN pipeline

Cascaded CNN pipelines have been successfully applied to various tasks in medical imaging [Ch17]. In this work a first CNN is trained to localize the aorta within the whole T2 abdominal scan (Fig. 3, left). This is necessary because there is a considerable variation between aorta positions of different slices. After the position has been found, the aorta region is cropped out of the whole input image and resized to 64x64. The resulting slices now mostly depict the actual aorta with little surrounding tissue, allowing the network to focus on segmenting the clinically important aorta region.
Next, a second CNN is used to find the final segmentation mask using the smaller cropped out aorta images of the first CNN as an input (Fig. 3, right). This specialization allows each of the two networks to focus on learning the appropriate weights for its specific task.

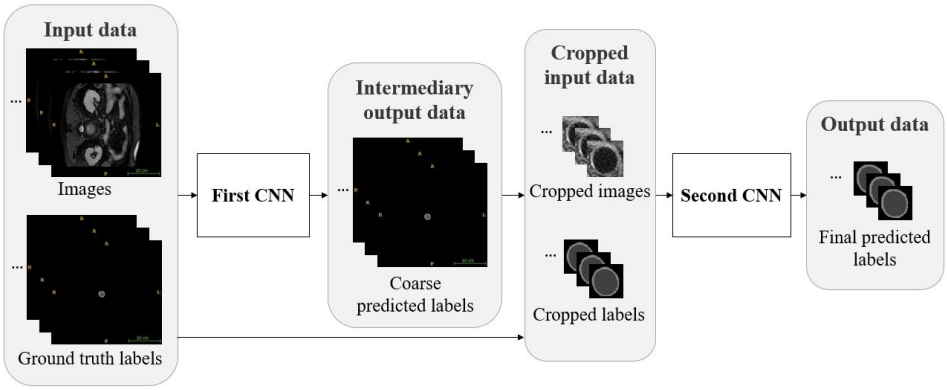


Fig. 3: Cascaded CNN pipeline

3.4 3D CNN

Similar to the previously depicted methods, 3D CNNs have also shown promising results in medical image segmentation [Ha17] [MNA16]. The access to additional information in the third dimension allows 3D CNNs to learn better, more discriminative features. The architecture used in this work is based on the 2D cascaded CNN pipeline of the last chapter (Fig. 3) with all 2D convolutional layers of the second CNN replaced by 3D ones. Also, in order to perform the memory-intensive 3D convolutions, the architecture was simplified by removing one level at the top of the architecture, resulting in a smaller overall depth of 4 levels, and by reducing the number of bottleneck blocks and channels in the other levels.

3.5 CNNs for slice classification

The second problem tackled in this work is to classify whole slices as containing plaque or not. Given the u-net inspired architecture of 3.2 a straightforward way of achieving a whole slice classification is to replace the prediction block with a fully connected layer followed by a softmax layer to get prediction probabilities for each of the two classes. This approach was the first method used in this work. As in the 3D architecture and all following classification methods, it is applied to the cropped out 64x64 aorta images and therefore acts as a second CNN in the cascaded pipeline. However, in other literature the upward path of the CNN is usually discarded for slice classification resulting in the second architecture used for plaque detection, which is depicted in Fig. 4.

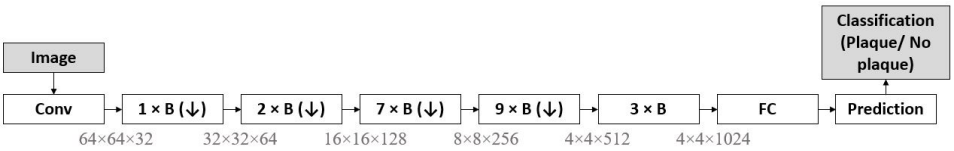


Fig. 4: CNN plaque slice classification architecture

Hereby, only the former downward contracting path is used and a fully connected layer followed by a softmax prediction layer are added to the end to obtain the class probabilities.

3.6 Capsule CNN

As a third classification method this paper adapts and applies the promising capsule networks to the more challenging medical imaging data. Similar to the original paper [SFH17], the architecture is less complex than comparable state-of-the-art CNNs and consists of two convolutional layers followed by a fully connected layer. The first convolutional layer has 256 filters with 9x9 kernels, stride 1 and is followed by a ReLU activation function. The second convolutional layer consists of 32 8-dimensional capsules and the final layer has a 16-dimensional capsule for each of the two output classes.

3.7 Quality measures

The quality of each presented segmentation method is evaluated via the Dice score metric [Di45] which measures the overlap of the predicted region with the ground truth region. It is defined in (1), where T is the ground truth region and P the predicted region:

$$dice = \frac{2 \times |P \cap T|}{|P| + |T|} \tag{1}$$

The Dice score assumes values between 0 and 1 with 0 equating to the worst and 1 to the best possible segmentation. The quality of the classification methods is analyzed using the metrics precision, accuracy and recall (2, 3, 4) [Po11].

$$precision = \frac{tp}{tp + fp} \tag{2}$$

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{3}$$

$$recall = \frac{tp}{tp + fn} \tag{4}$$

All three metrics are derived from the confusion matrix which gives the number of true positive (tp), true negative (np), false positive (fp) and false negative (fn) slices when comparing ground truth with predicted classification of plaque-containing slices.

4 Results

This section provides a short overview of the dataset, the setup of the experiments and the results of all 3 segmentation and all 3 slice-wise classification methods.

4.1 Dataset

The dataset consisted of 240 abdominal T2 scans acquired over a duration of several years. Each scan was composed of slices with the dimensionality of 512x512. A label mask associated with each scan that identified four different image regions, namely *background*, *blood-pool*, *aortic wall* and *plaque* was used as the ground truth. The dataset was randomly split into 50% training and 50% testing data to allow evaluation of the proposed methods. Hereby, the class imbalance between plaque-containing slices and non-plaque-containing slices in the overall dataset was maintained after the train-test split to retain realistic training and test conditions and enable results that generalize well.

4.2 Experimental setup

All experiments were run on a 12 GB Nvidia GPU with the tensorflow deep learning library. Adam was used as an optimizer for training. Experiments were run with three different loss functions, i.e. Dice loss, Jaccard loss and cross entropy loss. Cross entropy performed best

in all instances and was therefore used as a loss function in all reported results. The learning rate was set to 0.0005 and the batch size to 4 in all cases except the 3D model where it was set to 2 due to memory restrictions. Both values represent the optimum found after various experiments. The three aortic wall segmentation methods were evaluated via the standard Dice coefficient and the three classification methods via the metrics precision, accuracy and recall.

4.3 Aortic wall segmentation

Overall, all deep residual CNN architectures achieved good results in the aortic wall segmentation task. An example qualitative segmentation result is shown in Fig. 5. Both location and curvature of the aortic wall are captured very well by all CNN methods. Shortcomings can be seen at some portions of the aortic wall where the wall thickness of the network prediction appears slightly larger than in the corresponding ground truth image.

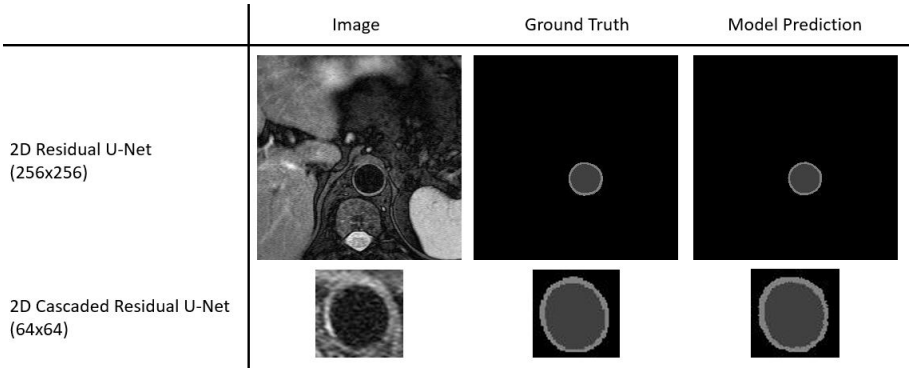


Fig. 5: Example of qualitative segmentation results

Quantitative results are presented in Tab. 1 with the u-net derived architecture of 3.2 in the first row. The next row refers to the cascaded architecture of 3.3 with a 2D CNN used as a second network while the last row corresponds to the cascaded CNN architecture with a 3D CNN used as a second CNN. Each of the three columns represents one of the aforementioned image regions.

Segmentation method	Background	Blood-pool	Aortic wall
Basic CNN	**0.999**	**0.958**	**0.797**
Cascaded 2D	0.991	**0.958**	0.791
Cascaded 3D	0.989	0.954	0.768

Tab. 1: Dice scores of different segmentation methods for various image regions

The *blood-pool* region was segmented very well by all examined methods with the basic CNN and the cascaded 2D method performing slightly better than the cascaded 3D technique.

The clinically most relevant region *aortic wall* was also segmented well by all methods with the basic CNN again achieving the highest score, closely followed by the cascaded 2D method.
Therefore the overall best performing method for aorta segmentation in this work was the basic CNN architecture, which achieved the highest Dice scores for all regions in question.

4.4 Plaque detection

Three different slice-wise plaque detection methods were analyzed using the precision, accuracy and recall metrics (Tab. 2). The first line in the table refers to the residual CNN without an upward path of Fig. 4. The second line depicts the results of the residual slice-wise classification u-net (3.5) while the third line indicates the capsule CNNs performance (3.6).

Classification Method	Precision	Accuracy	Recall
Residual CNN	**0.75**	0.61	0.25
Residual U-Net	0.66	0.67	0.27
Capsule CNN	0.74	**0.68**	**0.29**

Tab. 2: Precision, accuracy and recall of each of the three plaque slice classification methods

All methods achieved considerably better performance than random guessing. The residual CNN had the highest precision value with the capsule CNN only slightly behind. In both accuracy and recall the capsule CNN achieved the highest scores closely followed by the residual u-net in both cases. All in all, the best method for the binary classification of plaque slices in this work was therefore the capsule CNN, which had an almost as high precision score as the residual CNN and both the highest accuracy and recall scores.

5 Discussion

This section provides a discussion and potential explanations of the observed results.

5.1 Aortic wall segmentation

The basic CNN architecture performed best in segmenting all three different image regions. One major advantage of this method was that no resizing had to be performed to adjust the differently sized cropped aorta images to one resolution that the CNN could accept as input, which had to be done in both cascaded methods. This allowed the data to retain the highest degree of realness and not lose important textural information due to the inherent errors of resizing. Due to the aforementioned general complications of medical imaging data, this avoidance of further image degradation turned out to be decisive in the segmentation

task. Also, the larger 256x256 static ROI crop in the basic CNN allowed for a bigger area of the aorta's naturally surrounding tissue to remain within the image, which might have helped the CNN in better delineating the wall boundaries. This is in contrast to the original assumption that by cropping out as much surrounding tissue around the aorta as possible, the network would be able to focus on segmenting the important aorta region itself and not get distracted by other unimportant tissue. While this assumption might still hold, its gain was not enough to offset the errors induced by resizing.
In addition, considering multiple slices when performing a segmentation as in the 3D cascaded method worsened performance, indicating that there was a high variation in aortic wall characteristics between neighboring slices which confused the CNN instead of helping it learn from the additional data points of adjacent slices.

5.2 Plaque detection

The plaque detection task was best achieved by the capsule CNN. As compared to standard CNNs, capsule CNNs are better able to handle viewpoint changes of objects, i.e. to identify an object as the same object type even if the object itself was rotated or translated. Since plaque changes considerably in size, shape and position between individual slices, this property allows the capsule CNN to better detect its existence due to it storing a more complete representation of what constitutes plaque in the form of vectors as opposed to scalar values in the case of a standard CNN. Another advantage of capsule CNNs was their simpler architecture consisting of only three different layers, indicating a more efficient design and potentially leading to shorter training and running times. The residual CNN with no upward path performed considerably better on precision than the residual u-net because no information was lost in the transpose convolution operations of the upward path. While these operations are necessary in case of a full slice segmentation to obtain a mask with equal resolution to the input image, this work showed that they are detrimental when performing whole-slice classification.

6 Conclusion

This work examined various state-of-the-art methods for two problems in medical imaging. A u-net derived residual CNN with an increased number of layers and an organisation in bottleneck blocks achieved the best results in segmenting background, bloodpool and aortic wall in abdominal T2 images. Cascaded 2D and 3D approaches performed worse due to image quality loss in resizing and lack of surrounding tissue information due to narrower cropping around the aorta. The recently published capsule CNNs performed better than previous state-of-the-art methods in binary slice classification of challenging medical images due to their capacity to better represent and detect changes in plaque size, shape and position.
In the future, the promising but relatively simplistic capsule CNN architecture could achieve

even better results with a more elaborate and fine-tuned architecture design. Capsule CNNs could also be extended to perform full image segmentation instead of just classification. Furthermore, other recent advances in deep learning, such as atrous convolutions, might capture differences in shape, size and location in a better way due to their less static and more flexible kernels.

References

[Ar17] Araújo, T.; Aresta, G.; Castro, E.; Rouco, J.; Aguiar, P.; Eloy, C.; Polónia, A.; Campilho, A.: Classification of breast cancer histology images using Convolutional Neural Networks. PloS one 12/6, e0177544, 2017.

[Ch17] Christ, P. F.; Ettlinger, F.; Grün, F.; Elshaera, M. E. A.; Lipkova, J.; Schlecht, S.; Ahmaddy, F.; Tatavarty, S.; Bickel, M.; Bilic, P., et al.: Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:1702.05970/, 2017.

[Di45] Dice, L. R.: Measures of the amount of ecologic association between species. Ecology 26/3, pp. 297–302, 1945.

[Dr16] Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C.: The importance of skip connections in biomedical image segmentation. In: Deep Learning and Data Labeling for Medical Applications. Springer, pp. 179–187, 2016.

[Du12] Duquette, A. A.; Jodoin, P.-M.; Bouchot, O.; Lalande, A.: 3D segmentation of abdominal aorta from CT-scan and MR images. Computerized Medical Imaging and Graphics 36/4, pp. 294–303, 2012.

[Gu10] Gupta, S.; Berry, J. D.; Ayers, C. R.; Peshock, R. M.; Khera, A.; De Lemos, J. A.; Patel, P. C.; Markham, D. W.; Drazner, M. H.: Left ventricular hypertrophy, aortic wall thickness, and lifetime predicted risk of cardiovascular disease: the Dallas Heart Study. JACC: Cardiovascular Imaging 3/6, pp. 605–613, 2010.

[Ha17] Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.-M.; Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical image analysis 35/, pp. 18–31, 2017.

[He16] He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Pp. 770–778, 2016.

[Ho16] Hoffmann, U.; Massaro, J. M.; D'Agostino, R. B.; Kathiresan, S.; Fox, C. S.; O'Donnell, C. J.: Cardiovascular event prediction and risk reclassification by coronary, aortic, and valvular calcification in the Framingham Heart Study. Journal of the American Heart Association 5/2, e003144, 2016.

[KSH12] Krizhevsky, A.; Sutskever, I.; Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. Pp. 1097–1105, 2012.

[Li15] Liu, C.-Y.; Chen, D.; Bluemke, D. A.; Wu, C. O.; Teixido-Tura, G.; Chugh, A.; Vasu, S.; Lima, J. A.; Hundley, W. G.: Evolution of aortic wall thickness and stiffness with atherosclerosis: long-term follow up from the multi-ethnic study of atherosclerosis. Hypertension/, HYPERTENSIONAHA–114, 2015.

[Li16] Li, W.; Luo, S.; Luo, J.; Liu, Y.; Huang, W.; Chen, J.: Association between abdominal aortic plaque and coronary artery disease. Clinical interventions in aging 11/, p. 683, 2016.

[LSD15] Long, J.; Shelhamer, E.; Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Pp. 3431–3440, 2015.

[MNA16] Milletari, F.; Navab, N.; Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, pp. 565–571, 2016.

[Po11] Powers, D. M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation./, 2011.

[RFB15] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241, 2015.

[Ru15] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) 115/3, pp. 211–252, 2015.

[SFH17] Sabour, S.; Frosst, N.; Hinton, G. E.: Dynamic routing between capsules. In: Advances in Neural Information Processing Systems. Pp. 3859–3869, 2017.

[SZ14] Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556/, 2014.

[Wa17] Wang, Y.; Seguro, F.; Kao, E.; Zhang, Y.; Faraji, F.; Zhu, C.; Haraldsson, H.; Hope, M.; Saloner, D.; Liu, J.: Segmentation of lumen and outer wall of abdominal aortic aneurysms from 3D black-blood MRI with a registration based geodesic active contour model. Medical image analysis 40/, pp. 1–10, 2017.

[Wi17] Wilkins, E.; Wilson, L.; Wickramasinghe, K.; Bhatnagar, P.; Leal, J.; Luengo-Fernandez, R.; Burns, R.; Rayner, M.; Townsend, N.: European cardiovascular disease statistics 2017. European Heart Network: Brussels, Belgium/, 2017.