3D Pulmonary Artery Segmentation from CTA Scans Using Deep Learning with Realistic Data Augmentation

Karen López-Linares Román[1,2,4], **Isaac de La Bruere**[3], **Jorge Onieva**[4], **Lasse Andresen**[4], **Jakob Qvortrup Holsting**[4], **Farbod N. Rahaghi**[3], **Iván Macía**[1], **Miguel A. González Ballester**[2,5], **Raúl San José Estepar**[4]

[1]Vicomtech Foundation and Biodonostia, San Sebastián, Spain

[2]BCN Medtech, Universitat Pompeu Fabra, Barcelona, Spain

[3]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

[4]Applied Chest Imaging Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

[5]ICREA, Barcelona, Spain

Abstract

The characterization of the vasculature in the mediastinum, more specifically the pulmonary artery, is of vital importance for the evaluation of several pulmonary vascular diseases. Thus, the goal of this study is to automatically segment the pulmonary artery (PA) from computed tomography angiography images, which opens up the opportunity for more complex analysis of the evolution of the PA geometry in health and disease and can be used in complex fluid mechanics models or individualized medicine. For that purpose, a new 3D convolutional neural network architecture is proposed, which is trained on images coming from different patient cohorts. The network makes use a strong data augmentation paradigm based on realistic deformations generated by applying principal component analysis to the deformation fields obtained from the affine registration of several datasets. The network is validated on 91 datasets by comparing the automatic segmentations with semi-automatically delineated ground truths in terms of mean Dice and Jaccard coefficients and mean distance between surfaces, which yields values of 0.89, 0.80 and 1.25 mm, respectively. Finally, a comparison against a Unet architecture is also included.

Keywords

Pulmonary artery; Deep learning; CTA Convolutional neural network; Segmentation

---

klopez@vicomtech.org.

1. Introduction

The morphological assessment of the Pulmonary Artery (PA) is essential to evaluate several Pulmonary Vascular Diseases (PVD). Most patients with Pulmonary Hypertension (PH), present a remodeled main PA with a diameter considerably larger than that of a control subject and thus, being an important biomarker for predicting and detecting hypertension. In the Chronic Obstructive Pulmonary Disease (COPD), a widening of the PA is associated with increased risks of exacerbation and decreased survival rates. Pulmonary Embolism (PE) refers to the blockage of one of the pulmonary arteries, mostly caused by blood clots. Thus, it is essential to monitor the arterial obstruction to evaluate the severity of PE.

Computed tomography (CT) and CT angiography (CTA) play a crucial role in the diagnosis and management of PVD since they allow to assess macroscopic pulmonary vascular morphology quantitatively. In this study, we aim at lever-aging CTA images of several patient cohorts to segment the PA with a new 3D Convolutional Neural Network (CNN) architecture. Deep learning has already been applied to segment other vascular structures from CT images with promising results [3,7,10], which encouraged us to use it for PA segmentation.

2. Literature Review

The segmentation of PA can be challenging due to its complicated and variable shape, motion artifacts, and proximity to other blood vessels such as the pulmonary vein that may hamper the correct segmentation. Even if there are many studies in the literature about pulmonary vascular tree segmentation, they usually focus on vessel segmentation within the lungs or pulmonary emboli and nodule detection, without specifically analyzing the PA.

Regarding the segmentation of the PA outside the lung, which is our goal, only a few studies have been proposed. In [2] a Hessian matrix based preprocessing followed by a region growing method is proposed, which relies on a previous extraction of the lungs and the heart. The method in [14] also requires a priori knowledge of the artery morphology followed by a fast-marching algorithm and a registration to a target reference volume, which did not fully address the variability in PA sizes and shapes. In [6] a semi-automated tool which uses level sets and geodesic active contours to segment the main PA is presented, with the goal of measuring the PA diameters in patients with PH. From the obtained segmentations, the authors extract the artery centerline and measure the diameter, reporting a mean error up to 6 mm. A similar study to measure PA cross-sectional area is proposed in [9], where the artery is segmented using dilation and erosion operations on 14 normal patient CTA scans.

Compared to previous works in the literature, our method combines images from PE cohorts, PH cohorts, and control patients and is tested on many volumes. Additionally, it is fully automatic, it does not include any shape prior and it yields a mean error when measuring PA diameters of 2.5 mm.

3. Materials

A total of 51 CTA volumes of different patients are employed to train our CNN. Among these datasets, 39 patients have PE, 8 of them are control subjects who were thought to have PH, and the remaining 4 have hypertension. The mean intensity in the PA is higher than 550 HU in all the CTA volumes, and motion-related artifacts are present in most of the images. Figure 1 shows sample CTA slices of three patients coming from different cohorts.

To test the network, an additional 91 CTA volumes are used, all of them corresponding to patients with PE, being it our largest cohort. The mean intensity in the PA in these cases ranges between 350 HU and 550 HU.

3.1 Fuzzy Ground Truth Generation

For the 142 patients, ground truth labels are obtained semi-automatically using ITK-Snap [16]. The first step consists of selecting a region of interest around the PA, extracting a sub-volume that starts at the aortic valve and expands until the main PA is not observed.

Then, an initial segmentation is extracted with the region competition snake approach, using a thresholded version of the image as the feature image that drives the evolution and forces the snake to fit the boundary of the artery. The minimum and maximum thresholds employed to create the feature image for the training datasets are set to 500 HU and 900 HU, respectively, whereas for the test images, the employed thresholds are 300 HU and 900 HU. A seed point is placed within the main PA to initialize the evolution of the snake, which is manually stopped when an approximate segmentation is obtained. The parameters that control the evolution of the front, i.e. the region competition force and the smoothing or curvature force, are set to 1 and 0.5, respectively.

Finally, the output segmentation from the region competition snake approach is manually refined, as shown in Fig. 2. Two main corrections are applied:

- Removal of veins and other structures incorrectly labeled as arteries

- Inclusion of clots in the segmentation to ensure a natural artery shape

The resulting ground truth segmentations are considered fuzzy, since it is difficult to have a precise delimitation of the artery contour when there is a large clot in the artery. Additionally, small artery branches have not been consistently labeled across the different datasets.

4. Methods

Hereby, we propose a new 3D convolutional neural network for the segmentation of the PA from CTA volumes. The proposed network, fully described in Sect. 4.3, is inspired by the 3D V-net [8] with modifications introduced from the 2D Fully Convolutional DenseNet (FC-DenseNet) [5] and the 2D Efficient neural network (ENet) [11]. We employ a training strategy that relies on a strong use of data augmentation, mostly generated with realistic deformations, as explained in Sect. 4.1. Finally, we validate our network with the test set by comparing the semi-automatically generated ground truths with the network predictions in

terms of Dice and Jaccard scores. Since the final clinical goal is to characterize the aortic morphology, we also measure the distance at each point between the two surfaces, i.e., the ground truth and the output from the network.

4.1 Data Augmentation Using Realistic Deformations

Data augmentation have been largely used in deep learning in the biomedical field due to the limited number of annotated datasets. In particular, for the case of 3D datasets, it is difficult and time-consuming to obtain a corpus of annotated images that are large enough to account for the anatomical variability between subjects. Thus, researchers usually apply data augmentation techniques, mostly in the form of rotations and translations to generate new volumes. In [12] a new data augmentation approach was proposed, based on applying random elastic deformations to the original volumes. The use of these synthetically generated volumes seemed to be the key to train a segmentation network with very few annotated samples.

Inspired by this work, we efficiently augment our dataset using realistic elastic deformations as well as traditional rotations and translations. Unlike in [12], where the applied deformations were random, we propose to generate realistic deformation vectors from the Principal Component Analysis (PCA) of a subset of deformation fields extracted directly from the affine registration of several volumes. The steps are the following:

1. Register 10 CTA volumes to a reference volume of a control subject using 3D Slicer [1] and extract the 3D deformation fields corresponding only to the affine transformation

2. Extract the mean deformation and the eigenvectors and eigenvalues of the ten deformation fields using two PCA models:

    • PCA1-Model: considers the correlation between the components of the deformation fields, i.e., x, y, and z

    • PCA2 Model: considers each component of the fields independently

3. Generate new deformation fields by randomly weighing the first six eigenvectors (which account for most of the variability) with values from 0 to the square root of the corresponding eigenvalue

    • For PCA1-Model the three components are weighted equally

    • For PCA2-Model we weight x, y and z independently

4. Generate new synthetic volumes by applying these deformation fields to each original CTA volume in the training set, as shown in Eq. 1 for PCA1-Model and in Eq. 2 for PCA2-Model.

$$\tilde{I}_j : \sum_{i=1}^{6} <w_i * B_i> + \mu \tag{1}$$

$$\tilde{I}_j: \sum_{i=1}^{6} < w_{x_i} * B_{x_i} > + \mu_x + \sum_{i=1}^{6} < w_{y_i} * B_{y_i} > + \mu_y$$

$$+ \sum_{i=1}^{6} < w_{z_i} * B_{z_i} > + \mu_z$$

(2)

where $\tilde{I}_j$ is the generated synthetic image, w_i are the weights generated from the eigenvalues, B_i are the eigenvectors, and μ is the mean image extracted from the 10 original deformation fields.

Following this procedure, we create 50 new volumes per each original input CTA. 30 of them are extracted with PCA1-Model, whereas another 20 are generated with PCA2-Model. This allows the network to learn invariance to deformations without the need to see these transformations in the annotated image corpus. This is particularly important in biomedical segmentation since deformation is the most common variation in tissue and realistic deformations can be simulated efficiently with the proposed approach. Examples of the generated volumes in 2D and 3D are shown in Figs. 3 and 4, respectively.

4.2 Related Networks Served as Inspiration

The **V-Net** [8] network is one of the few architectures in the literature specifically designed to work with 3D images. It is composed of convolution, deconvolution and pooling layers arranged in an encoding and a decoding path. Every couple of layers in the encoding path a down-convolution is performed, and for every pooling the number of feature maps is doubled to allow the network to distribute the information from the previous layer throughout the maps, instead of losing it when reducing the spatial resolution. Before each down-convolution, a skip-connection is introduced to pass higher resolution maps to the decoding path. In the decoding path, an up-convolution is performed every couple of layers and feature fusion with the skip connections is applied, improving the convergence time and the quality of the segmentation.

FC-DenseNet [5] is one of the most recent networks for 2D semantic segmentation. As the V-Net, FC-DenseNet also uses an encoding and a decoding pathway to obtain global features, incorporating feature fusion. However, opposed to the idea in V-Net, this architecture uses many convolutional layers but each of them with few channels, whereas in V-net there are fewer convolutional layers and the information is distributed in more filters. Each layer is directly connected to every other layer in a feed-forward fashion and batch normalization is implemented before all convolutional layers, which helps to control over-fitting.

Finally, in [11] the **ENet** is proposed, which aims at providing real-time semantic segmentation by using a low amount of parameters, squeezing in as much information as possible in every parameter. A critical contribution of ENet is the introduction of a down-sampling block that combines max pooling and strided convolution to avoid representational bottlenecks.

4.3 Proposed Convolutional Neural Network

Figure 5 shows the main building blocks of our proposed network, displayed in Fig. 6. It has an encoding and a decoding path as the V-Net and the FC-DenseNet. As in FC-DenseNet, the input is propagated through the network via dense connections and channels are appended throughout. The structure of the encoder is also changed to an ENet style block. We also remove some layers as compared to FC-DenseNet, but increase the width. The number of filters in each regular dense block is increased gradually. In the decoding pathway, we decrease the number of channels steadily to reach an amount that is computationally feasible without performing extreme information compression.

The network is implemented using Keras with tensorflow. It is trained with 3468 volumes extracted by augmenting the scans of 91 different patients. All volumes are resized to $128 \times 128 \times 64$ and the intensities are rescaled to 0–1.

The model is built in a Xeon E7 3.6 GHz, 62 GB processor equipped with a Nvidia GeForce GTX1080 card, under Linux Ubuntu 16.04 SMP 64 bits. We train the network using ADAM optimization with a batch size of 1, an initial learning rate of 1e–03 and plateau learning rate decay with a factor of 0.2 when the validation loss is not improved after five epochs, with a minimum learning rate of 1e–05. We use the binary accuracy metric and try to minimize the binary cross entropy loss function. Early stopping is also applied to avoid overfitting, thus, stopping the learning process after 20 epochs, as shown in Fig. 7.

Finally, the model is tested on the 91 less contrasted CTA scans described in Sect. 3. The predictions are 3D probability maps where the intensity of each pixel is the probability of it being PA. We apply gaussian smoothing to the output grayscale image, followed by Otsu's thresholding that aims at selecting an optimal case-specific threshold when the image contains two classes following bi-modal histogram and voting binary hole filling to obtain the final binary segmentation.

4.4 Validation Approach

To evaluate the performance of our network, we compare the automatically obtained segmentation with the fuzzy ground truths in terms of Dice and Jaccard scores for the 91 test cases, and we calculate the mean and standard deviation.

Since the final clinical goal is to characterize PA morphology, i.e. its diameter, we generate the 3D surfaces of both segmentations using VTK [13] to calculate the mean distance between them. First, we use the Discrete Marching Cubes method to extract the surfaces and the normals at every point. Then, we create a Kd-tree spatial decomposition of the set of points of each surface. Finally, we use a point locator to find the closest point in the ground truth surface for every point in our segmentation, and we measure the Euclidean distance between them. The distance between surfaces is the mean distance of all the points in the surface, which corresponds to the mean error when measuring the PA radius.

5 Results

Table 1 summarizes the results for the proposed network using realistic deformation-based data augmentation and without using it. Our method yields a mean Dice coefficient of 89% and a Jaccard score of 80%. From the clinical point of view, when measuring the PA radius our method falls into a mean error of 1.25 mm. According to several studies [4,15], the PA diameter of a control subject is smaller than 29 mm and in patients with PH the artery is enlarged. Hence, the mean error made with our segmentation approach falls at least below 8.6%.

Figure 8 depicts the box plots for the validation scores for all the patients used for testing, where some clear outliers that negatively impact the achieved mean values are observed. The most noticeable two cases correspond to patients with a very big liver, in which the network gets confused and segments part of the liver as if it were the artery (see Fig. 10). Our guess is that the network may interpret that this region corresponds to the end part of the artery branch.

Regarding the use of deformation-based data augmentation, an improvement of 2.3% and 3.9% is obtained for Dice and Jaccard coefficients, respectively. For the distance between surfaces, an improvement of 1.57% is achieved. As shown in Fig. 8, the Dice and Jaccard score's improvement is statistically significant according to the Wilcoxon test but it is not for the distance between surfaces.

Finally, we also trained and tested the V-net in [8] to compare the results, which are shown in Table 2. Even if the Dice and Jaccard scores are very similar for both architectures, the distance between surfaces is much larger in the case of the Unet and the statistical significance is notable, with a p-value of 1.73e−09 for the case of the distance according to the Wilkoxon test (Fig. 9). This suggest that our architecture enables better quantification of mean pulmonary artery diameters.

6 Conclusions

Hereby, we proposed a new CNN to PA segmentation from CTA images, which opens up the opportunity for more complex analysis of the evolution of the PA geometry (i.e. going beyond just measuring the diameter). The network is based on an encoder-decoder scheme similar to the V-net [8], but by including Dense blocks and Enet blocks, we are able to improve the segmentation results, mostly in terms of distance between surfaces. Adding bootstrapping to the loss function could further increase the accuracy of our model.

Additionally, a novel data augmentation approach has been described, which relies on a PCA analysis of deformation fields extracted from the affine registration of several volumes. For the current work, 10 different base deformation fields have been extracted by registering 10 volumes to a reference CTA. Looking at the results, it seems that more fields are necessary to account for a larger anatomical variability between patients since the improvement as compared to training without this data augmentation is not statistically significant regarding the distance between surfaces. However, a tendency is observed in the Dice and Jaccard scores, which suggests that with more deformation fields a better outcome

may be achieved. Additionally, the fields generated to create the synthetic images after the PCA analysis are obtained by varying the weight of each eigenvector with the square root of the corresponding eigenvalue, which limits the range of deviation from the mean deformation. Weighting each eigenvector with a wider value range could also account for more variability in the input data.

Finally, regarding future work, our aim is to incorporate a data augmentation technique that simulates non-contrast CT volumes from CTA scans. This may allow to use the same network to segment and characterize the artery in cohorts where the use of contrast is not usual, such as COPD patients.

Acknowledgments

References

1. 3D Slicer. https://www.slicer.org/

2. Ebrahimdoost Y, Qanadli SD, Nikravanshalmani A, Ellis TJ, Shojaee ZF, Dehmeshki J: Automatic segmentation of pulmonary artery (PA) in 3D pulmonary CTA images. In: Proceedings of the DSP, pp. 1–5 (2011)

3. Ibragimov B, Toesca D, Chang D, Koong A, Xing L: Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning. Phys. Med. Biol 62(23), 8943–8958 (2017) [PubMed: 28994665]

4. Collins J, Stern EJ: Chest Radiology, the Essentials. Lippincott Williams & Wilkins, Philadelphia (2007)

5. Jégou S, Drozdzal M, Vázquez D, Romero A, Bengio Y: The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. CoRR abs/1611.09326 (2016)

6. Linguraru MG, Pura JA, Uitert RLV, Mukherjee N, Summers RM: Segmentation and quantification of pulmonary artery for noninvasive CT assessment of sickle cell secondary pulmonary hypertension. Med. Phys 37(4), 1522–1532 (2010) [PubMed: 20443473]

7. Meijs M, Manniesing R: Artery and vein segmentation of the cerebral vasculature in 4D CT using a 3D fully convolutional neural network. In: Proceedings of the SPIE, vol. 10575, p. 6 (2018)

8. Milletari F, Navab N, Ahmadi S: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 3DV, pp. 565–571 (2016)

9. Moses D, Sammut C, Zrimec T: Automatic segmentation and analysis of the main pulmonary artery on standard post-contrast CT studies using iterative erosion and dilation. Int. J. Comput. Assist. Radiol. Surg 11(3), 381–395 (2016) [PubMed: 26410842]

10. Nardelli P, Jimenez-Carretero D, Bermejo-Peláez D, Ledesma-Carbayo MJ, Rahaghi FN, Estépar RSJ: Deep-learning strategy for pulmonary artery-vein classification of non-contrast ct images. In: Proceedings of the ISBI, pp. 384–387 (2017)

11. Paszke A, Chaurasia A, Kim S, Culurciello E: ENet: a deep neural network architecture for real-time semantic segmentation. CoRR abs/1606.02147 (2016)

12. Ronneberger O, Fischer P, Brox T: U-Net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015)

13. Schroeder W, Martin K, Lorensen B: The Visualization Toolkit–An Object-Oriented Approach to 3D Graphics, 4th edn. Kitware Inc., Clifton Park (2006)

14. Sebbe R, Gosselin B, Coche E: Segmentation of opacified thorax vessels using model-driven active contour. In: Proceedings of the IEEE EMBS, vol. 3, pp. 2535–2538 (2005)

15. Truong QA, et al.: Reference values for normal pulmonary artery dimensions by noncontrast cardiac computed tomography: the Framingham heart study. Circ. Cardiovasc. Imaging 5(1), 147–154 (2012) [PubMed: 22178898]

16. Yushkevich PA, et al.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006) [PubMed: 16545965]
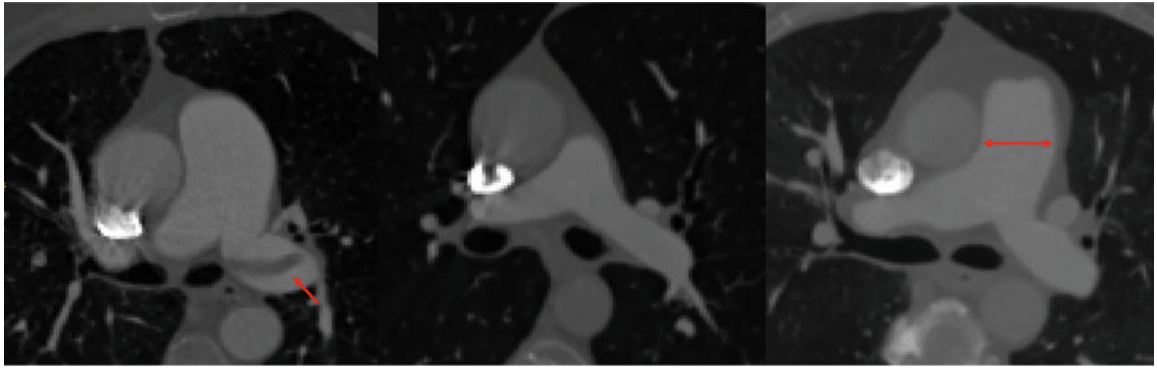
Fig. 1.
Sample CTA slices of 3 patients from different cohorts. Left - Pulmonary embolism dataset, where the arrow points towards a clot; Middle - Control subject; Right - Pulmonary hypertension case, where the arrows show a dilated artery.
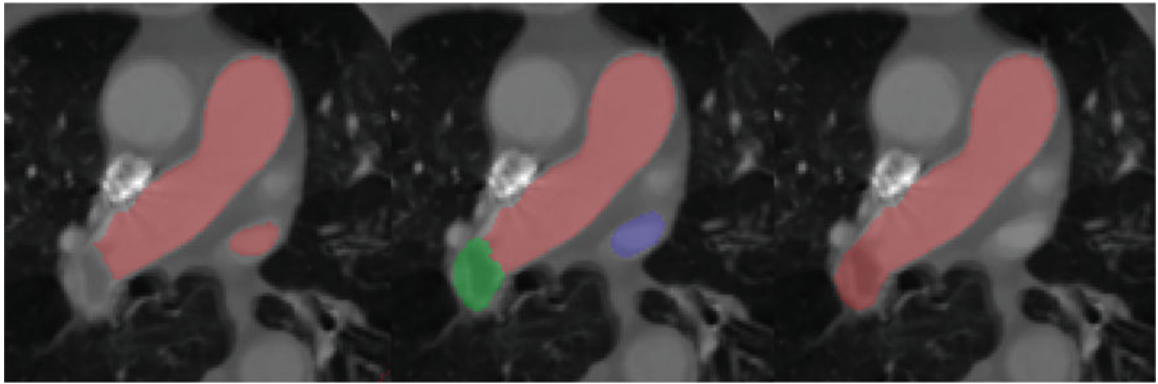
Fig. 2.
Correction of the automatically generated ground truth labels. Left - Automatically obtained segmentation; Middle - Correction of the segmentation by including the clot (green) and removing the vein (blue); Right - Final fuzzy ground truth used for the CNN.
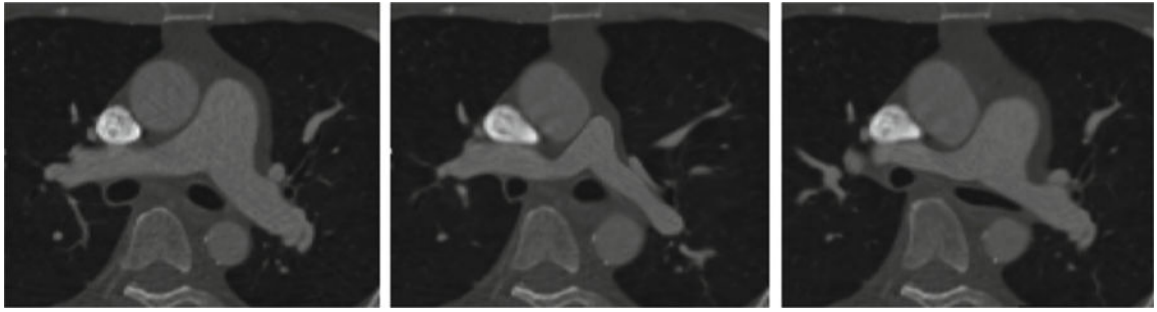
Fig. 3.
Sample axial slices of volumes generated using the realistic deformation based data augmentation technique. Right: original axial slice; Middle: corresponding slice generated using PCA2-Model; Left: corresponding slice generated using PCA1-Model.
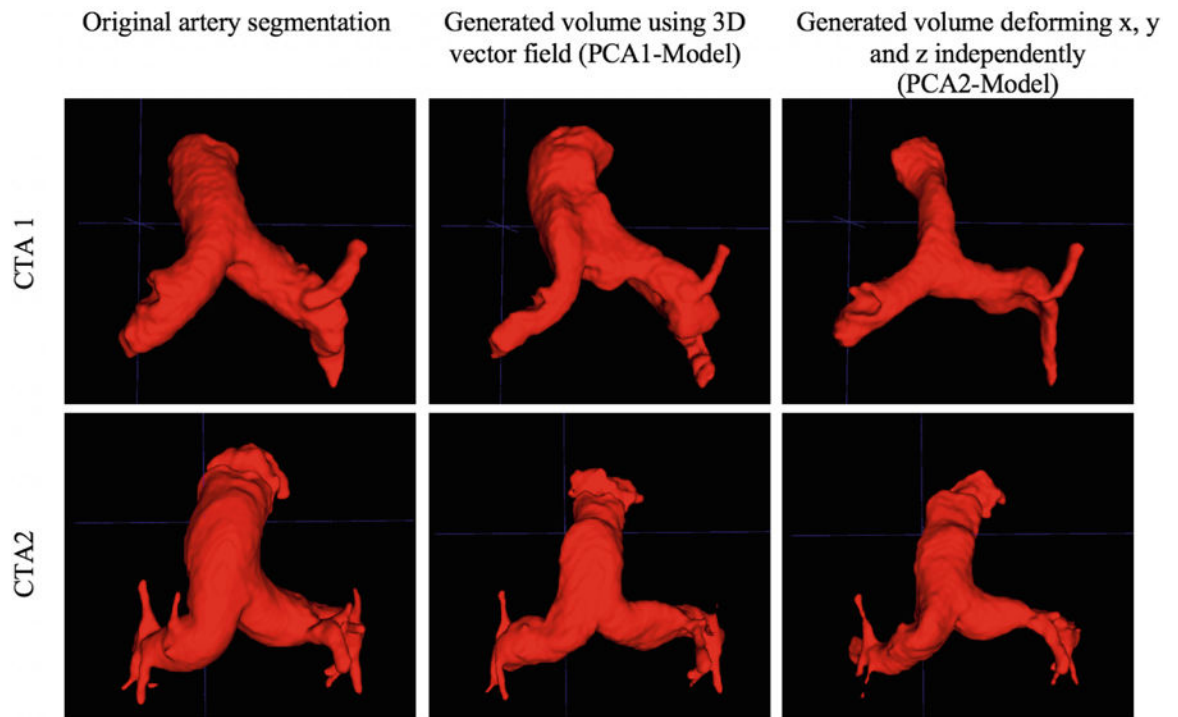
Fig. 4.
Sample volumes generated using the realistic deformation based data augmentation technique.
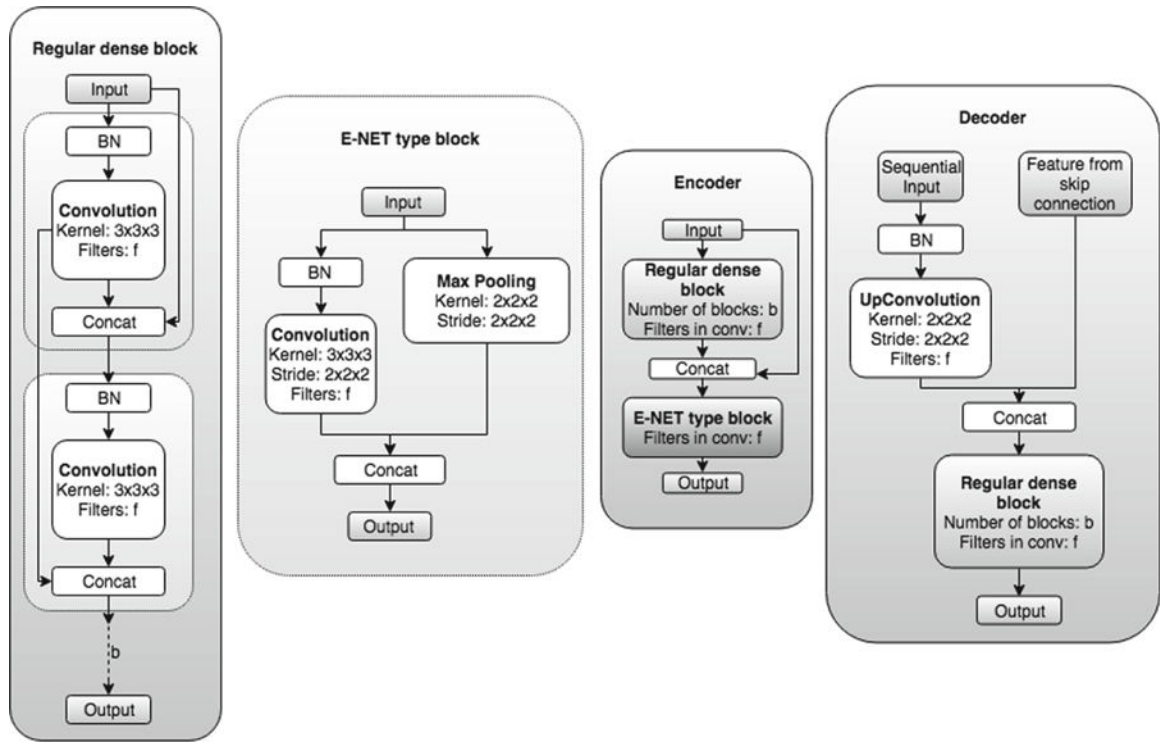
Fig. 5.
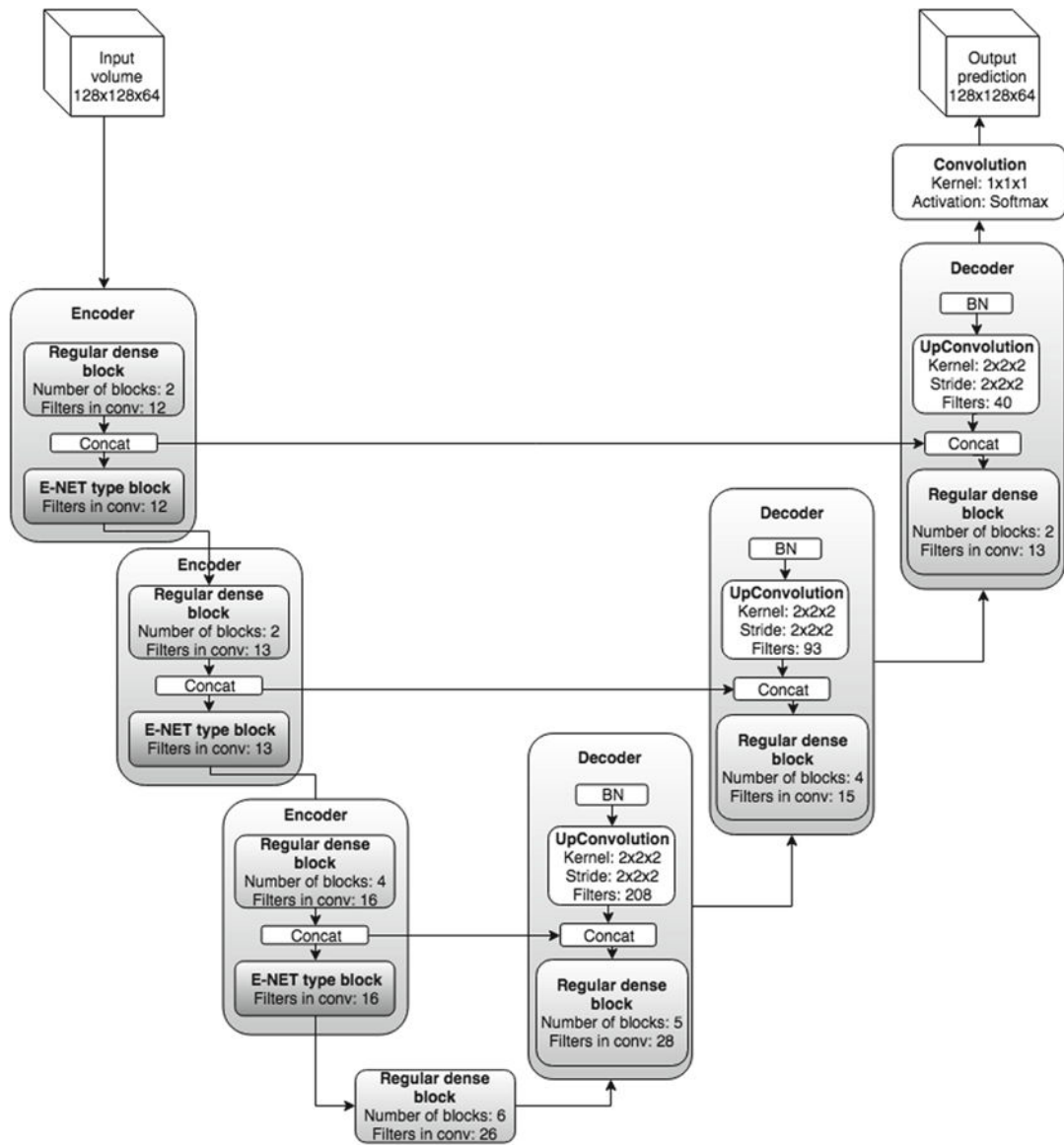The several blocks that compose the proposed convolutional neural network.

Fig. 6.
Scheme of the proposed convolutional neural network.

Fig. 7.
Training and validation loss and accuracy curves and fitted polynomial trendline as a function of epochs. Over-fitting is observed after epoch 20.
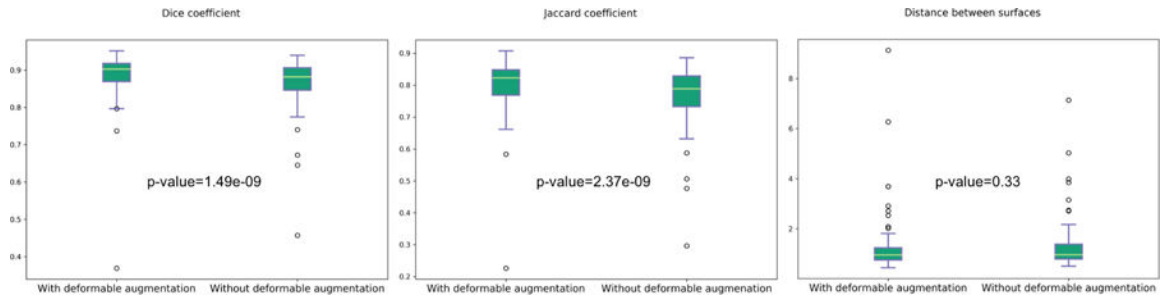
Fig. 8.
Plots showing the Dice and Jaccard scores and the mean distance between surfaces for all the test volumes when using the proposed data augmentation technique, and without it. The p-values corresponding to the Wilkoxon test are also displayed.
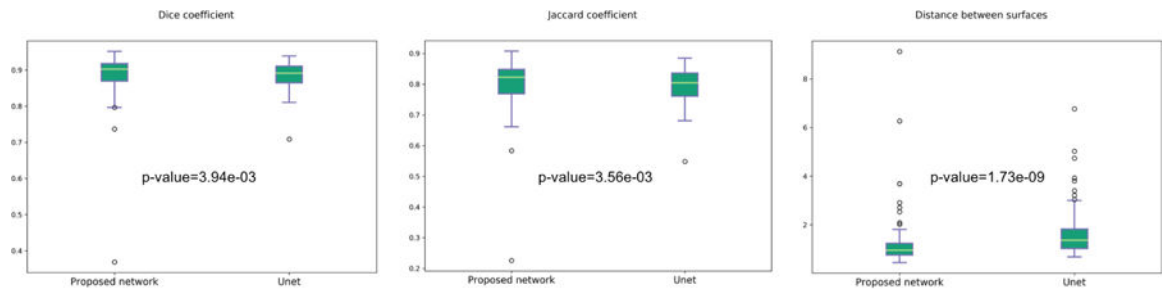
Fig. 9.
Plots showing the Dice and Jaccard scores and the mean distance between surfaces for the proposed architecture and a Unet. The p-values corresponding to the Wilkoxon test are also displayed.
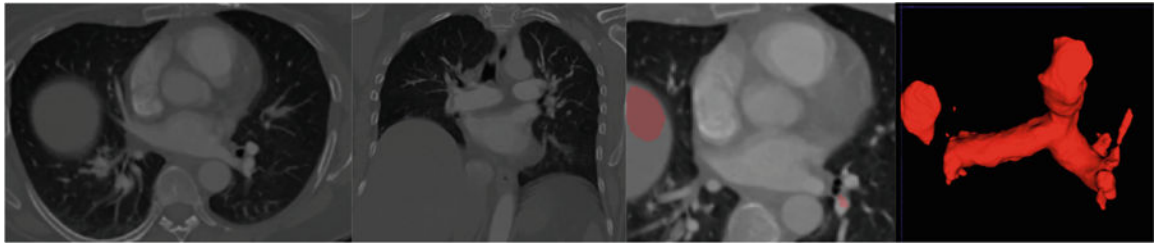
Fig. 10.
Outlier test case of a patient with a very big liver, which the network segments as artery.

Table 1.

Evaluation metrics for the proposed network when including realistic deformable registration based data augmentation and without it.

	Mean Dice Score	**Mean Jaccard Score**	**Mean distance between surfaces (mm)**
Without augmentation	0.87 ± 0.07	0.77 ± 0.09	1.27 ± 0.98
With data augmentation	0.89 ± 0.07	0.80 ± 0.09	1.25 ± 1.17

Table 2.

Evaluation metrics for the proposed method as compared to a traditional Unet when using the deformation-based data augmentation.

	Mean Dice Score	Mean Jaccard Score	Mean distance between surfaces (mm)
Unet	0.89 ± 0.04	0.80 ± 0.05	1.66 ± 1.03
Proposed architecture	0.89 ± 0.07	0.80 ± 0.09	1.25 ± 1.17