Extraction of Aortic Knuckle Contour in Chest Radiographs Using Deep Learning

Zhiyun Xue, Rodney Long, Stefan Jaeger, Les Folio, George R. Thoma, and Sameer Antani

***Abstract*— In this paper, we aim to extract the aortic knuckle (AK) contour in chest radiographs, an anatomical structure rarely being addressed in the literature. Since the AK structure is small and thin, simply adopting the deep network methods that are successful for large organ segmentation is inadequate for achieving good pixel-level accuracy and resolving local ambiguities. To address this challenge, we propose a new coarse-to-fine segmentation approach which focuses on global and local information contexts, respectively. Two convolutional networks are used. For the coarse segmentation, we use Faster-RCNN; for the fine segmentation, we use U-Net. Our evaluation uses the publicly available JSRT dataset; the results are quite promising. Besides presenting these results, we analyze issues such as the imprecision of manual contour marking, and automatic generation of the coarse segmentation ground-truth mask used for deep network training. Our approach is general and can be applied to extract other curve-like objects-of-interest.**

I. Introduction

Chest radiographs are frequently used by clinicians to diagnose diseases and conditions in and around the thorax. Automatic analysis of chest radiographs has been an active research field for decades. One of the main research topics is segmentation or extraction the region of interest. The majority of the work focuses on identifying regions of lung, heart, clavicle, and rib. Our group also has developed algorithms to segment these major regions [1, 2]. In this paper, we present our work on extracting aortic knuckle, a task that is rarely being addressed in the literature. Aortic knuckle (AK), also called aortic knob, is a radiographic structure that represents the part of the thoracic aorta arching backwards over the left main bronchus and pulmonary vessels. As shown in Figure 1, it refers to the hump-shape contour of aorta seen in frontal radiographs on the mediastinal silhouette. Abnormal aortic knuckle shape, such as the sign of enlargement and aortic nipple, may indicate cardiovascular disease [3].

Deep learning has become one of the most popular research topics in computer vision due to several converging factors, including its state-of-the-art performance, in particular for classification tasks, the open-source code contributed by the community, the affordability of GPU hardware to make network training time practical, and the access to large annotated datasets. Although most of the work concentrates on classification because of its significant success, deep learning has been investigated for other research tasks and achieved very promising performance. For semantic segmentation, the cornerstone deep network is the fully convolutional network (FCN) proposed in 2014 [4]. It is the first CNN architecture that can be trained end-to-end and generate segmentation maps for input images of arbitrary size. Stemming from FCN, subsequent important deep networks for object segmentation include encoder-decoder networks, and multiscale networks that incorporate context knowledge [5, 6]. For reviews on recent deep learning techniques developed for semantic segmentation, please refer to [7]. Most of these segmentation deep networks were evaluated on general-domain image datasets.

Deep learning techniques have been applied for analyzing chest x-rays; the majority are for classification (such as classifying images into certain types of clinical manifestations [8]). There are only a few works on using deep learning to segment regions (lungs, heart, and clavicles) in chest radiographs (with lungs being the main focus) [9, 10]. The work on extracting aortic knuckle/arch is very scarce. The only publications we found are the two patents [11, 12], in which the aortic knuckle/arch is extracted based on selecting "arc candidates" from the lung boundary. In addition, no performance evaluation of their methods is provided in these two patents. Compared to the regions of lung, heart, clavicle and ribs, aortic knuckle segmentation faces several key challenges. These challenges include: 1) the size/length of aortic knuckle is much smaller compared to the size of the whole image; 2) as aortic knuckle is a thin open contour on the boundary of the mediastinum, it is labor-intensive to delineate it precisely. Therefore the method should be robust and not sensitive to the imprecision of manual marking; 3) shape and intensity alone are not sufficient to identify the aortic knuckle; spatial context information is crucial. To address these challenges, we propose a new coarse-to-fine deep learning approach to use and balance local and global context information. To make the method robust, we also employ pre-processing and post-processing procedures. The proposed method is tested on the Japanese Society of Radiological Technology (JSRT) dataset, and the results are quite promising. The general framework can be applied to extract other contour landmarks, such as contours of the hemidiaphragms and costophrenic (CP) angles.

II. Method

As pointed out in Section I, unlike the extraction of large organ regions, the aortic knuckle has several intrinsic characteristics that require special consideration. The aortic knuckle, which shown in chest radiographs as a small part of the mediastinum silhouette boundary contour, contains significantly fewer pixels than other anatomical landmarks and the whole chest image. Although the shape of aortic knuckle looks generally like an arch-like curve, the shape

All the authors are with the National Library of Medicine, Bethesda, MD 20894 USA (corresponding author: Zhiyun Xue; phone: 301-827-4939; e-mail: xuez@mail.nih.gov).

alone is not adequate to identify it in the image; this requires both global as well as local context information. We propose a coarse-to-fine deep learning segmentation method for this task. Our method consists of two main steps. In the first step, we extract the coarse region that includes the aortic knuckle. This step uses additional context information surrounding the aortic knuckle which is more recognizable and easier for the network to learn. In the second step, we identify the aortic knuckle inside that coarse region-of-interest (ROI). Since the first step removes unwanted area and reduces the search space for the second step significantly, the aortic knuckle contour becomes much more prominent in the coarse ROI and it is easier for the network to learn to differentiate it from other regions based on its shape, intensity and location. The convolutional networks that are used for object detection and semantic segmentation can be adopted for the coarse and fine segmentation steps, respectively. Specifically, in this paper, the Faster-RCNN [13], a two-stage object detection CNN network, is employed in the coarse segmentation step, and U-Net [14], a semantic segmentation FCN architecture that has been successfully applied to biomedical images, is used in the fine segmentation step.

In the following, we introduce the image data, the generation of the ground truth masks especially for the step of coarse segmentation, the Faster-RCNN and the U-Net architectures, and the pre- and post- processing methods.

A. Image data

We use the JSRT dataset [15] for both training and testing. The JSRT dataset consists of 247 posterior-anterior chest x-ray images. The dataset originally aimed for nodule detection and among the 247 images, 154 have nodules, and 93 do not have a nodule. This dataset has been used to evaluate segmentation of left/right lung fields, heart, and left/right clavicles. We use this dataset for evaluating our method of aortic knuckle contour extraction.

B. Ground truth masks generation

For both steps (coarse segmentation and fine segmentation), we use convolutional networks to do the segmentation. Therefore, ground truth markings are needed in both steps to train the network.

1) AK contour manual drawing

Since it would be time-consuming and tedious for a radiologist to manually draw the AK contours on all the images, we adopted the method of "engineer draw with expert radiologist guidance". We manually drew them, after radiologist consultation, then modified them based on feedback from the radiologist. We used the path tool in GIMP to do the boundary drawing. GIMP output SVG files, which we converted to PNG.

2) Ground truth ROI mask for coarse segmentation

For coarse segmentation, at the beginning, we manually created the ROI boxes using the Matlab Training Image Labeler tool. We required that a ROI mask for coarse segmentation meets the following criteria: it is a rectangular box around the mediastinum that includes the aortic knuckle; it should cover part of the right/left lung and there should have some gaps between the aortic arch, and the box boundaries. These criteria aim to: 1) include the context information that are more recognizable and easier for the network to learn; 2) be relaxed enough to allow easy drawing. Figure 2 (a) and 2(b) show an example of the aortic knuckle contour and ROI markings, respectively. We later developed a method to automatically generate these ROI masks, in order to reduce the cost for ground truth labeling especially when a large number of images need to be annotated. As illustrated in Figure 3, the method is based on the location of the manual markings of the aortic knuckle and the location of the mediastinum center line. Specifically, the right/upper/lower boundary of the ROI box are decided by adding a gap from the corresponding the right/upper/lower boundary of the bounding box of the aortic knuckle contour respectively; and the left boundary of the ROI box is decided by the distance from the center line to the right boundary of the ROI box. The mediastinum center line is extracted by finding the location on the x-axis that corresponds to the peak of the vertical profile of the central part of the image (for example, cropped ¼ of the original border length on each side, in order to remove the outer area that may also have high vertical profile value). Both the manual or automatic generated ROI box is made to be a square ROI box (centered at the center of the ROI box with the side length equal to the maximum of the width and length of the ROI box).

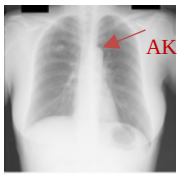

Figure 1. Illustration of aortic knuckle.

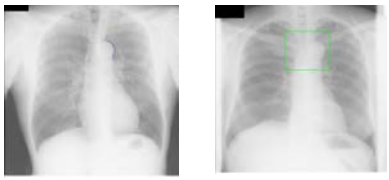

Figure 2. Examples of manual AK contour markings and ROI markings

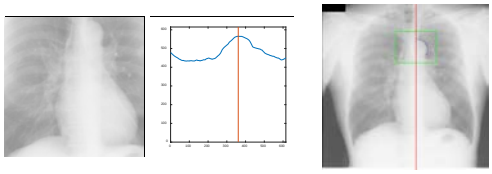

Figure 3. Illustration for aortic knuckle ROI mask generation.

Please note that there exists imprecision in the ground truth data. For the fine segmentation step (i.e., manual marking of the AK contour), it is due to the difficulty of precise contour drawing. For the coarse segmentation step (i.e., manual or automatic generation of the ground truth ROI mask), it is due to the relaxed definition of the ROI. The techniques we developed for tolerating this imprecision will be described in the section of pre-processing and post-processing.

C. Segmentation convolutional network

1) Faster-RCNN

Faster-RCNN is an object detection algorithm based on a unified network of two networks: a Region Proposal Network (RPN) and a Fast-RCNN [16] object detection network. The RPN is used to output a set of bounding box proposals for an input image. The Fast-RCNN is used to classify each of the bounding box proposals generated by RPN to be one of the objects or the background as well as refining the bounding boxes of the objects. The RPN and Fast-RCNN share a common set of convolutional layers (base convolutional network). The RPN slides a small spatial window on the

feature map output by the last convolutional layer and estimates the probability of being an object or background (*objectness score*) for each of the bounding boxes predefined at each location with different scales and aspect ratios (called *anchors*) as well as adjusting the coordinates of these anchors. The bounding box proposals output by RPN are ranked based on their objectness scores and are post-processed to remove highly overlapped ones using non-maximum suppression (NMS). The top-N ranked region proposals are then input to Fast-RCNN. The Fast-RCNN contains a ROI pooling layer which extracts a fixed-length feature vector from the feature map for each of the input region proposal. The feature vector, after going through a couple of fully connected layers, is then input to two sibling output layers: one outputs the estimates for M +1 classes (M object classes plus the background); the other outputs 4 values representing the bounding-box position for each of the object classes. Faster-RCNN can be trained by training the RPN and Fast-RCNN either alternately or jointly. For the detailed description on the algorithm and implementation of Faster-RCNN, please refer to [13]. For our application, we use Faster-RCNN as a single-object detector to identify the AK ROI. The software package we use is [17]. We use the VGG16 network pre-trained using ImageNet data as the base convolutional network. The number of training iterations is set to be 10k.

2) U-Net

U-Net is a semantic segmentation convolutional network proposed by [14] which works with limited training images. It has an encoder-decoder architecture which consists of a contracting path and a symmetric expanding path. The contracting path contains pairs of convolutional layers and pooling layers to capture the context, while the expanding path recovers the spatial dimension and adds object details by merging the extracted features from the contracting path. For the details of the U-Net architecture, please refer to [14]. U-Net has been applied/adapted to segment objects in biomedical images such as cells, retina vessels, and lungs and has achieved state-of-the-art performance. In this paper, we adapt this network to segment the aortic knuckle contour in the fine segmentation step. The specific network architecture is given in Figure 4 (where Conv$K$-N indicates a convolutional layer with N feature maps of filter size $K\times K$). It contains convolutional, drop-out, max pooling and upsampling layers. The dropout rate is 0.2 for all the drop-out layers. The pooling size is 2×2 for both the max pooling layers and the upsampling layers. The exponential linear units (ELUs) are used as the activation function for all the convolutional layers except the last convolutional layer; the binary cross entropy based loss function is used and the network is trained using the stochastic gradient descent (SGD) optimizer with a learning rate of 10^{-2}.

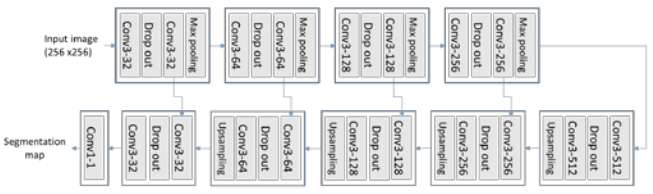

Figure 4. U-Net architecture

For both steps, we use the 154 images with nodules as the training data and the 93 images without nodules as the test data. Data augmentation is not applied in the coarse segmentation step, but is applied in the fine segmentation step in order to generate more training data for performance improvement.

D. Preprocessing and post-processing

1) For the step of coarse segmentation

In this step, no pre-processing is applied to the input image to the Faster-RCNN except that the images are resized to 600 x 600, zero-centered with mean-subtraction, and normalized with the standard deviation. For the test images, after getting the output bounding box of AK ROI from the trained Faster-RCNN, we also generate the corresponding AK ROI square box. Figure 5 shows three examples of the Faster-RCNN testing results (the ground truth ROI box, the ROI box output by Faster-RCNN, and the corresponding square ROI box are indicated in green, blue and red color respectively). The test images cropped by using the AK ROI square box are then resized to 256 x 256 and fed to the U-Net for the fine segmentation.

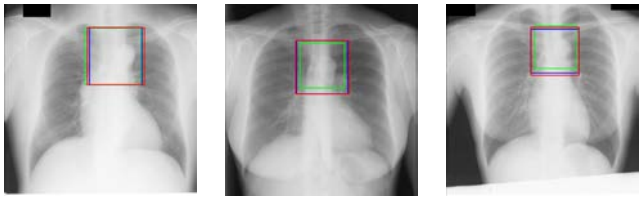

Figure 5. Faster-RCNN results.

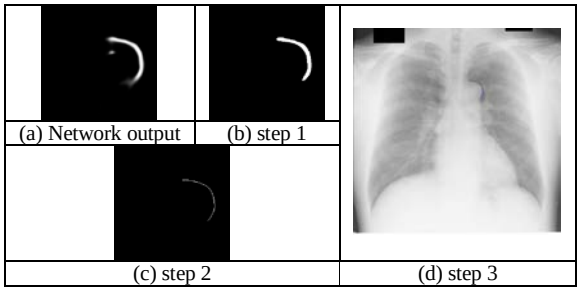

Figure 6. Post-processing of AK contour map in fine segmentation

2) For the step of fine segmentation

In this step, the U-Net is trained with the pairs of the cropped images and cropped AK contour masks generated based on the AK ROI masks. The AK contour masks used by the U-Net are first dilated. The reasons for applying dilation to make the contour mask thicker are as follows: 1) the single-pixel width aortic knuckle contour drawn by hand may not be precisely located along the edge of the mediastinum silhouette; 2) the dilated contour may cover more area of the edge that exhibits intensity transition which makes it easier for the network to learn to represent. In this step, we did data augmentation in order to improve segmentation performance. It was done by shifting the center of the AK ROI mask in each of the 4 directions and using different cropping sizes when cropping. The cropped images/masks are then contrast enhanced using histogram equalization, resampled to 256x256, zero-centered with mean-subtraction, and normalized with the standard deviation. The mean and standard deviation are calculated from the whole training set.

These cropped images generated/augmented from the original 154 images in the training set are used to train the U-Net. As shown in Figure 6, given a test input image, the output segmentation probability map of the U-Net is post-processed as follows to get the final contour: 1) convert the map to a binary image by thresholding and extract the largest connected-component as the object of interest; 2) extract the center line of the object as the AK contour by thinning; 3) resize the segmentation probability map and add the outside area (being removed when cropping) back so that the size of the map is the same as the original image.

III. Experimental Results

For the coarse segmentation, the number of training images is 154 and the number of test images is 93. To evaluate the coarse segmentation performance, we check that the extracted ROI region encloses the whole AK contour. This is very important since the goal of the coarse segmentation step is to reduce the search space for the next step. The subsequent segmentation of AK contour will fail if the ROI region output from the coarse segmentation does not include it. For each of the test images, the entire ground truth AK contour is within the extracted ROI region. We also use the Dice coefficient to measure the difference between the extracted AK ROI regions and ground truth. The average Dice coefficient is 0.88 for the test set.

For the fine segmentation, we augment the data; the number of training images is 13860. The training images are then split 90/10% for training/validation sets when training the U-Net. The input binary mask to the U-Net is dilated with a disk-shaped structuring element of radius being 4 pixels, as described in Section II.D.2. For the test set, we use Dice coefficient to measure the performance of the segmentation result of the U-Net after the first post-processing step. We also use the Pratt's figure of merit (PFOM) to compare the extracted contour (after thinning) with the ground truth contour, as [18] indicates that PFOM stood out among contour detection evaluation criteria. The PFOM is defined as:

$$PFOM = \frac{1}{\max(I_{gt}, I_c)} \sum_{i=1}^{I_c} \frac{1}{1+\alpha d^2(i)} \quad (1)$$

where I_{gt} is the number of pixels in the ground truth contour, I_c is the number of pixels in the segmented contour,α is the scaling constant, and $d(i)$ is the distance between the i^{th} pixel belonging to the segmented contour I_c and the nearest pixel in the reference contour I_{gt}. For the 93 test images, the average Dice coefficient is 0.62 and the average value of PFOM is 0.59. We also evaluate the performance on the test image set that is augmented using the same way as the training images and the number of images is 8370. The average Dice coefficient is 0.63 and the average value of PFOM is 0.60 for the augmented test set. Figure 7 shows several examples of AK contour extraction results (blue: extracted contour, green: ground truth contour).

IV. Conclusion

In conclusion, we present a method for extracting the aortic knuckle contour in chest radiographs, an anatomical structure rarely being targeted in the literature. There are three key contributions of our work: 1) we carry out the first study on segmenting AK contour with deep learning techniques; 2) we propose a coarse-to-fine framework to address the distinctive challenges facing aortic knuckle segmentation for which simply adopting the deep networks used for large organs segmentation is not sufficient; 3) we provide techniques to deal with the imprecision of the manual markings or relaxed ground truth masks. We tested this method on the JSRT dataset; the results are promising. To improve the performance, especially for the fine segmentation, future work includes exploring other state-of-the-art deep networks, improving the manual AK contours, and collecting more data. The proposed approach is quite general and can be easily generalized to segment other curve-like objects contained in regions which can be coarsely characterized by landmark features, for example, the cardio-diaphragmatic angles.

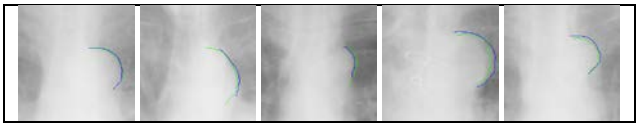

Figure 7. Examples of AK contour extraction results

Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and Lister Hill National Center for Biomedical Communications.

References

[1] S. Candemir, S. Jaeger, K. Palaniappan, J. Musco, R.K. Singh, Z. Xue, A. Karargyris, S.K. Antani, G.R. Thoma, C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imaging*, vol33, no 2, pp.577-590, Feb. 2014.

[2] S. Candemir, S. Jaeger, S.K. Antani, U. Bagci, L.R. Folio, Z. Xu, G.R. Thoma, "Atlas-based rib-bone detection in chest X-rays," *Comput. Med. Imaging Graph.*, vol.51, pp.32-39, Jul. 2016.

[3] S. Beslic, N. Beslic, S. Beslic, A. Sofic, M. Ibralic, J. Karovic, "Diagnostic imaging of traumatic pseudoaneurysm of the thoracic aorta," *Radiology and Oncology*, vol. 44, no. 3, pp 158-163, 2010.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," arXiv preprint, arXiv:1511.00561, 2015.

[6] L.C. Chen*, G. Papandreou*, I. Kokkinos, K. Murphy, and A. L. Yuille (*equal contribution), "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs", arXiv preprint, arXiv:1606.00915, 2016.

[7] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv preprint, arXiv:1704.06857, 2017.

[8] P. Lakhani, B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary Tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[9] A. Kalinovsky, and V. Kovalev, "Lung image segmentation using deep learning methods and convolutional neural networks," In: *XIII Int. Conf. on Pattern Recognition and Information Processing*, 3-5 October, Minsk, Belarus State University, 2016, pp. 21-24.

[10] A. A. Novikov, D. Lenis, D. Major, J. Hladůvka, M. Wimmer, K. Bühler, "Fully convolutional architectures for multi-class segmentation in chest radiographs," arXiv preprint, arXiv: 1701.08816, 2017.

[11] Z. Huo, J. Zhang, "Method for detecting anatomical structures," US Patent 8,577,108. 2013.

[12] Z. Huo, M. Xu, J. Zhang, "Method for detecting anatomical structures," US Patent App. 12/190,616. 2010.

[13] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems* (NIPS), 2015.

[14] O. Ronneberger, P. Fischer, T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention* (MICCAI), Springer, LNCS, vol.9351, pp. 234-241, 2015.

[15] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *AJR*, vol. 174, pp. 71-74, 2000.

[16] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision* (ICCV), 2015.

[17] https://github.com/YuwenXiong/py-R-FCN.

[18] S. Chabrier, H. Laurent, C. Rosenberger, B. Emile, “Comparative study of contour detection evaluation criteria based on dissimilarity measures,” *EURASIP Journal on Image and Video Processing*, 2008.