

Segmentation of Multiple Structures in Chest Radiographs Using Multi-task Fully Convolutional Networks

Chunliang Wang^(✉)

School of Technology and Health (STH),
KTH Royal Institute of Technology, Stockholm, Sweden
chunliang.wang@sth.kth.se

Abstract. Segmentation of various structures from the chest radiograph is often performed as an initial step in computer-aided diagnosis/detection (CAD) systems. In this study, we implemented a multi-task fully convolutional network (FCN) to simultaneously segment multiple anatomical structures, namely the lung fields, the heart, and the clavicles, in standard posterior-anterior chest radiographs. This is done by adding multiple fully connected output nodes on top of a single FCN and using different objective functions for different structures, rather than training multiple FCNs or using a single FCN with a combined objective function for multiple classes. In our preliminary experiments, we found that the proposed multi-task FCN can not only reduce the training and running time compared to treating the multi-structure segmentation problems separately, but also help the deep neural network to converge faster and deliver better segmentation results on some challenging structures, like the clavicle. The proposed method was tested on a public database of 247 posterior–anterior chest radiograph and achieved comparable or higher accuracy on most of the structures when compared with the state-of-the-art segmentation methods.

Keywords: Multi-task deep neural network · Fully convolutional network · Image segmentation

1 Introduction

Chest radiography is one of the most common medical imaging procedures for screening and diagnosis of pulmonary diseases thanks to its low radiation and cost. Although the interpretation of chest radiography is deemed as a basic skill of a certified radiologist, the inter- and intra-observer performance is highly variable due to the subjective nature of the reviewing process [1]. To assist in the diagnosis of chest radiography, a number of computer-aided diagnosis/detection (CAD) systems have been developed to provide a second opinion on a selective set of possible pathological changes, such as lung nodules [1] or tuberculosis [2]. In such a system, segmentation of various structures from the chest radiograph is often performed as an initial step. The accuracy of the segmentation often has a strong influence on the performance of the following steps such as lung nodule detection, or cardiothoracic ratio quantification, therefore a robust and accurate lung field and heart segmentation method is essential for these systems. A number of

dedicated segmentation methods have been proposed to target some specific organs, most commonly the lungs in chest radiographs [3–6]. However, the segmentation problem is still a great challenge, even for human observers, due to large variation of anatomical shape and appearance, inadequate boundary contrast and inconsistent overlapping between multiple organs (e.g. the relative positions between bones, muscles and mediastinum) which are hard to be modeled with the common statistical models. In [7, 8], we proposed a hierarchical-shape-model guided multi-organ segmentation method for CT images, that outperformed some dedicated single organ segmentation methods. This gave some support to the hypothesis that solving multiple organ segmentation simultaneously may be better than solving a single organ segmentation problem as the algorithm gets more context information. In this study, we extended the same philosophy to the chest radiography segmentation with a totally different segmentation framework, based on convolutional neural networks (CNN).

CNN, or deep neural networks in general, has gained popularity in recent years due to the outstanding performance on a number of challenging image analysis problems, such as image classification, object detection and semantic segmentation [9], as well as a variety of medical applications [10]. In contrast to conventional machine learning approaches that use handcrafted features designed by a human observer, CNNs automatically adjust the weights of convolutional kernels to create data-driven features that optimize the learning objectives at the end of the neural network. Recently, an increasing number of reports have suggested that adding multiple objectives or combining features trained for different objectives, so-called multi-task CNNs, can deliver better results than the single-task models [11, 12]. In this study, we implemented a multi-task fully convolutional network (FCN) to simultaneously segment multiple anatomical structures, namely the lung fields, the heart, and the clavicles, in standard posterior-anterior chest radiographs. When tested on a public database of chest x-ray images, the proposed method achieved comparable or higher accuracy on most of the structures than the state-of-the-art segmentation methods.

2 Method

2.1 Fully Convolutional Network

In general, a CNN consists of a number of convolutional layers followed by a number of fully connected layers. This setup requires the input images/image patches to share the same size. When used for image segmentation, it requires the input image to be converted to a series of largely overlapping patches around each pixel. This makes the computation very inefficient. FCN can be seen as an extension of the classical CNN, where the fully connected layers are removed or replaced by convolutional layers [9]. This allows FCNs to be applied to images of any size and output label maps proportional to the input image. Combined with “skips” and up-sampling or deconvolution layers [9], the output map can have the same size as the input image. This design eliminates redundant computation on overlapping patches and makes both the training and testing processes more efficient than the patch-based CNN approaches. In this study, we implemented a variation of FCN, called U-Net, which was proposed by

Ronneberger et al. [13]. The overall architecture of the U-Net used in this study is illustrated in Fig. 1. The left arm of the “U” shape consists of four repeating steps of convolution and max pooling. In the convolution steps, we used two consecutive 3×3 convolutional kernels. The exact number of features at each layer is given in the figure. At each max pooling step, the feature maps are reduced to half the size. The right arm of the “U” shape consists of four repeating steps of up-sampling and convolution, which allows the network to output a segmentation mask of the same size as the input image. Right after each up-sampling, the feature maps from the corresponding layers on the left arm are merged with the up-sampled feature maps before the following convolution operations. This allows the network to combine the context information

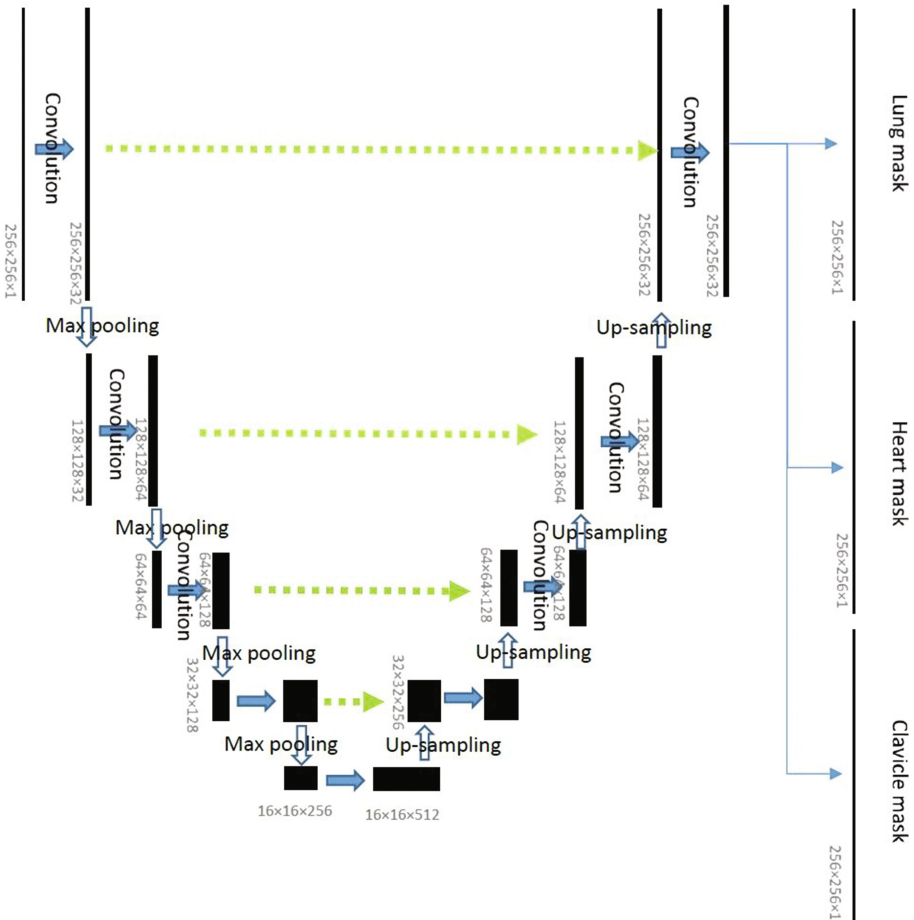


Fig. 1. The U-Net architecture used in this study. Black lines represent the images and feature maps (The height and width of the lines symbolize the size of the maps and the number of features at each step respectively, numbers on the side indicate the exact dimension of the features maps). Blue arrows represent convolutional operations. Dashed green arrows represent the ‘skips’ (Color figure online)

from the coarse layer and the detailed image features at the finer scale. The final segmentation masks are usually generated with a convolution layer with kernel size of 1×1 that combines the multiple feature maps with a softmax or sigmoid activation function. Compared with the original FCN method reported in [9], U-Net does not require the up-sampling layers to be trained from coarse to fine in multiple stages, but to train all layers in a single round, therefore it is easier to use in practice.

2.2 Multiple Single-Class FCN Vs. Multi-class FCN Vs. Multi-task FCN

In this study we try to segment the lungs, heart and clavicles from a chest x-ray image (cf. Fig. 2). There are a number of options to achieve multiple-class segmentation with FCN. One of the most trivial approaches is to treat the multi-structures separately and train multiple FCNs to segment different structures one after another. Another relatively straightforward method is to simply treat the structure labels as multi-classes and train a single FCN that outputs multiple probability maps with a single objective function. However, as shown in Fig. 2, the segmentation labels of the clavicles overlap largely with the lung fields. The softmax activation function that is commonly used to output multi-class predictions requires the samples to exclusively belong to a single class, and therefore cannot be directly used in this task. Multi-task FCN is another option which can generate multiple segmentation masks through a shared base FCN. The general principle of a multi-task deep neural network is to hybrid multiple output paths with different objective functions on top of a common base network (e.g. Fig. 1). It can also be interpreted as multiple output paths sharing the same feature pool that can be used for multiple tasks. While training a network that optimizes two tasks, such as landmark detection and segmentation, the feature pool may eventually contain two sets of features that complement each other. In the FCN setup, low-resolution layers are also thought of as context information for the finer output layers. In multi-task FCN, the context information is also enriched. In this study, we added 5 output nodes to the U-Net output layer that are expected to generate segmentation masks for 5 different structures. Each node is associated with its own loss function. In this study, we used the negative Dice coefficient as the loss function for each individual region, which eliminates the need of tuning the class weighting factors in situations where there is a strong imbalance between the number of foreground and background voxels [14], like the clavicles. The weighting factors for different objective functions were all set to 1.0, i.e. all structures are equally important.

2.3 Post-processing

As shown in Fig. 2, the output from FCN sometimes developed holes inside the targeted structure or islands outside. To remove these artifacts, we applied a fast level set method [15] that first shrinks from the border of the image with a relatively high curvature force (curvature weighting set to 0.7) and then expands with a low curvature force (curvature weighting set to 0.3). Because the output from U-Net is a probability map that is in a range from 0 to 1, we simply used 0.5 as the threshold to generate the

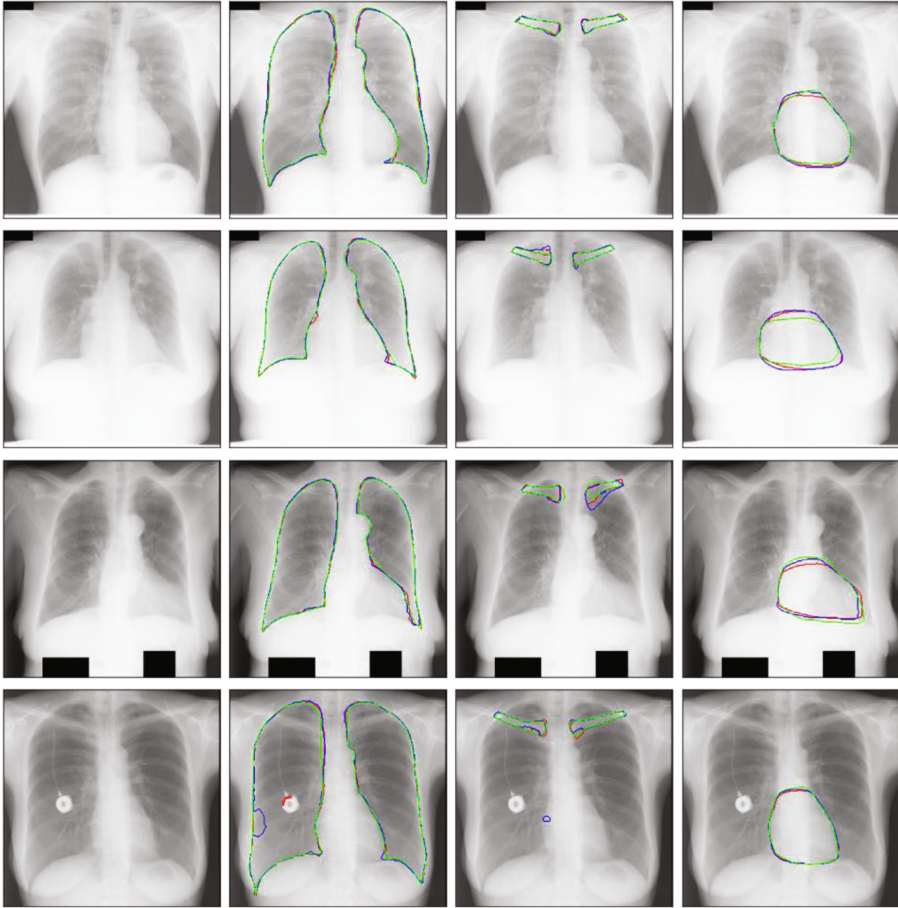


Fig. 2. Four representative cases. Columns from left to right are: input images, segmentation of lungs, segmentation of the clavicle and segmentation of the heart. Green contours represent the ground truth, blue contours represent the segmentation results from the single-task FCN, and red contours represent the segmentation results from the multi-task FCN (Color figure online)

speed-map for the level set method. The shrinking and expanding are similar to mathematical morphology opening and closing operations, but we have found the level set method to be more robust against large holes caused by unexpected objects in the images such as pacemakers or central venous access devices.

3 Experiments and Results

To validate the proposed method, we tested the proposed multi-task FCN approach on the public JSRT dataset of 247 chest radiographs with the segmentation masks from [3]. The original image size was 2048×2048 , but was down-sampled through linear

interpolation to 256×256 as suggested in [3]. Our implementation was based on the Keras framework with Theano backend (<http://keras.io>). The Adam optimizer with a learning rate of 0.0001 was used. The optimization was stopped at 50 epochs. A single-class FCN and a multi-class FCN were also implemented for comparison. For single-class FCN, five different FCNs were trained for five different structures with the negative Dice coefficient as the loss function. The same optimizer and learning rates were used for the lung segmentation network, but for training the clavicle and heart segmentation networks, a lower learning rate (1×10^{-5}) was needed to make the loss function to decrease and the number epoch was manually determined while monitoring the loss function (around 150 epochs for the heart and 300 epochs for the clavicles). For multi-class FCN, we used a sigmoid function as the activation function, and categorical cross entropy as the objective function. Similar optimizer settings as the multi-task FCN were used. Fivefold cross validation was used to generate the overlap measurements (Jaccard index). In each fold around 200 images were used for training. Using data augmentation technique, we generated 2000 random rotated and scaled image and mask pairs for the actual training. Table 1 compares the segmentation accuracy of three different methods without the post-processing step. In Table 2, the final results from the proposed method after post-processing are shown and compared with state-of-art results found in literature. In addition to the overlap measurement, the average distance (AD) between the manual contour and the segmentation result is also given for comparison.

Table 1. Multiple structure segmentation results using different FCN architectures

Structure name	Jaccard index of single-class FCN	Jaccard index of multi-class FCN	Jaccard index of multi-task FCN	P-value*
Left lung	0.957 ± 0.027	0.951 ± 0.041	0.955 ± 0.026	0.486
Right lung	0.963 ± 0.017	0.957 ± 0.046	0.962 ± 0.015	0.250
Left clavicle	0.845 ± 0.062	0.491 ± 0.171	0.863 ± 0.057	0.001
Right clavicle	0.838 ± 0.061	0.532 ± 0.156	0.862 ± 0.055	<0.001
Heart	0.899 ± 0.044	0.855 ± 0.093	0.898 ± 0.046	0.789

*: P value of student t-test between single-class FCN and multi-task FCN

Table 2. Quantitative comparison of the segmentation accuracy of the proposed method with other state-of-art methods reported on the same dataset

Methods	Lung field		Heart		Clavicle	
	Jaccard	AD (mm)	Jaccard	AD (mm)	Jaccard	AD (mm)
Observer [3]	0.946 ± 0.044	1.64 ± 0.69	0.878 ± 0.054	3.78 ± 1.82	0.896 ± 0.037	0.68 ± 0.26
Ginneken et al. [3]	0.949 ± 0.020	1.62 ± 0.66	0.860 ± 0.056	4.24 ± 1.87	0.736 ± 0.106	1.88 ± 0.93
Shao et al. [4]	0.946 ± 0.018	1.70 ± 0.76	–	–	–	–
Ibragimov et al. [5]	0.953 ± 0.020	1.43 ± 0.85	–	–	–	–
Hogeweg et al. [6]	–	–	–	–	$0.860 \pm 0.100^*$	$1.09 \pm 1.57^*$
Multi-task FCN	0.959 ± 0.017	1.29 ± 0.80	0.899 ± 0.044	3.12 ± 1.80	0.863 ± 0.045	0.98 ± 0.58

*: The results were reported on a different dataset.

4 Discussion and Conclusion

Chest radiograph segmentation is relative challenging for most conventional methods that are based on handcrafted features, mostly due to its complicated texture pattern. CNNs or FCNs, on the other hand, can cope with the complicate pattern more easily through the data-driven feature composition. As shown in Table 1, for the larger structures, such as the lungs and heart, both single-task and multi-task FCN delivered very promising results that outperform most existing methods summarized in Table 2. The results suggest that the segmentation variability of FCNs is even smaller than the inter-observer variability. However, the segmentation of the clavicles proved to be more challenging due to its size and complex surrounding structures (clavicles overlap with the lung, rib cage, vertebral column and sternum, as well as the soft tissue in the mediastinum). In our experiments, we found that multi-task FCN that segments multiple structures simultaneously, not only reduces the training and running time, but also helps the deep neural network to converge faster and deliver better segmentation results on the clavicles than the single-class FCN trained on single structure masks. One possible explanation is that the context information of the lungs helped the network to determine the boundary of the clavicles. Also the image features that are learned for lung segmentation can be used for the clavicle segmentation, which allows us to use a higher learning rate (10 times higher) when training the multi-task FCN than training a single-class FCN on clavicle alone. These findings are in line with the finding of some other multi-organ segmentation studies [7, 8] and multi-task CNNs studies [11, 12].

It is important to point out that the clavicle segmentation masks used in this study contain only those parts superimposed on the lungs and the rib cage have been indicated as shown in Fig. 2. The reason for this, as explained in [3], is that the peripheral parts of the clavicles are not always visible on a chest radiograph. Due to the overlapping between the clavicle and lung masks, the conventional softmax activation function is inapplicable in this case. In our experiments, we found that a sigmoid activation function combined with the categorical cross entropy objective function gives the best results on all five structures in the multi-class FCN setup. Other objective functions, including negative Dice score, were also tested, but failed to deliver better results. As shown in Table 1, the segmentation accuracy of the multi-class FCN on the clavicles is much worse than the other two methods.

In [10], the author also adapted a multi-task CNN framework for medical image segmentation. However, in their work, they trained a network to segment different structure from different image modalities. No comparison of segmentation accuracy between single organ and multi-organ segmentation was made.

In conclusion, we found that FCN-based image segmentation outperformed most conventional methods on lung field, heart and clavicle segmentation in chest radiograph. Multi-task FCN seems to be able to deliver better results on the more challenging structures. Our future works include to test the proposed method on a large dataset and extend it to handle 3D structure segmentation in CT or MRI volumes.

References

1. Kakeda, S., Moriya, J., Sato, H., Aoki, T., Watanabe, H., Nakata, H., Oda, N., Katsuragawa, S., Yamamoto, K., Doi, K.: Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *Am. J. Roentgenol.* **182**, 505–510 (2004)
2. Melendez, J., Sánchez, C.I., Philipsen, R.H.H.M., Maduskar, P., Dawson, R., Theron, G., Dheda, K., van Ginneken, B.: An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci. Rep.* **6**, 25265 (2016)
3. van Ginneken, B., Stegmann, M.B., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Med. Image Anal.* **10**, 19–40 (2006)
4. Shao, Y., Gao, Y., Guo, Y., Shi, Y., Yang, X., Shen, D.: Hierarchical lung field segmentation with joint shape and appearance sparse learning. *IEEE Trans. Med. Imaging* **33**, 1761–1780 (2014)
5. Ibragimov, B., Likar, B., Pernuš, F., Vrtovec, T.: Accurate landmark-based segmentation by incorporating landmark misdetections. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1072–1075. IEEE (2016)
6. Hogeweg, L., Sánchez, C.I., de Jong, P.A., Maduskar, P., van Ginneken, B.: Clavicle segmentation in chest radiographs. *Med. Image Anal.* **16**, 1490–1502 (2012)
7. Wang, C., Smedby, Ö.: Multi-organ segmentation using shape model guided local phase analysis. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 149–156. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_18](https://doi.org/10.1007/978-3-319-24574-4_18)
8. Jimenez-Del-Toro, O., Muller, H., Krenn, M., Gruenberg, K., Taha, A.A., Winterstein, M., Eggel, I., Foncubierta-Rodriguez, A., Goksel, O., Jakab, A., Kontokotsios, G., Langs, G., Menze, B., Salas Fernandez, T., Schaer, R., Walleyo, A., Weber, M.-A., Dicente Cid, Y., Gass, T., Heinrich, M., Jia, F., Kahl, F., Kechichian, R., Mai, D., Spanier, A., Vincent, G., Wang, C., Wyeth, D., Hanbury, A.: Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. *IEEE Trans. Med. Imaging* **35**, 2459–2475 (2016)
9. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **3**, 640–651 (2017)
10. Moeskops, P., Wolterink, J.M., Velden, B.H.M., van der Gilhuijs, K.G.A., Leiner, T., Viergever, M.A., Išgum, I.: Deep learning for multi-task medical image segmentation in multiple modalities. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 478–486. Springer, Cham (2016). doi:[10.1007/978-3-319-46723-8_55](https://doi.org/10.1007/978-3-319-46723-8_55)
11. Yu, B., Lane, I.: Multi-task deep learning for image understanding. In: 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 37–42. IEEE (2014)
12. Li, X., Zhao, L., Wei, L., Yang, M.-H., Wu, F., Zhuang, Y., Ling, H., Wang, J.: DeepSaliency: multi-task deep neural network model for salient object detection. *IEEE Trans. Image Process.* **25**, 3919–3930 (2016)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016)
15. Wang, C., Frimmel, H., Smedby, O.: Fast level-set based image segmentation using coherent propagation. *Med. Phys.* **41**, 73501 (2014)