A Multi-Task Convolutional Neural Network for Renal Tumor Segmentation and Classification Using Multi-Phasic CT Images — Source link ↗

Tan Pan, Huazhong Shu, Jean-Louis Coatrieux, Guanyu Yang ...+7 more authors

Institutions: Southeast University, University of Rennes, Nanjing Medical University

Related papers:

- Training on Polar Image Transformations Improves Biomedical Image Segmentation

- Multi-loss convolutional networks for gland analysis in microscopy

- Segmentation of Pulmonary CT Image by Using Convolutional Neural Network Based on Membership Function

- Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM

- Nested Dilation Network (NDN) for Multi-Task Medical Image Segmentation

A Multi-Task Convolutional Neural Network for Renal Tumor Segmentation and Classification Using Multi-Phasic CT Images

Tan Pan, Guanyu Yang, Chuanxia Wang, Ziwei Lu, Zhongwen Zhou, Youyong Kong, Lijun Tang, Xiaomei Zhu, Jean-Louis Dillenseger, Huazhong Shu, et al.

A multi-task convolutional neural network for renal tumor segmentation and classification using multi-phasic CT images

Tan Pan and Guanyu Yang and Chuanxia Wang and Ziwei Lu and Zhongwen Zhou and Youyong Kong
Lijun Tang and Xiaomei Zhu and Jean-Louis Dillenseger and Huazhong Shu and Jean-Louis Coatrieux

Abstract— Accounting for nearly 2% of all adults, renal cell carcinomas are sensitive to laparoscopic partial nephrectomy (LPN) which needs an accurate diagnosis and localization before operation. Faced with various intensity distribution, erratic location, irregular shape, etc, the image classification and semantic segmentation on CT scans of renal tumor are challenges. This paper presents a multi-task network, segmentation and classification convolutional neural network (SCNet), for preoperative assessment of renal tumor. Via the combination of two tasks, semantic features are fed to the classification network and classification results give segmentation network feedbacks in return. Besides, a 2-step segmentation strategy is conducted to the segmentation module which improves the result by 2.8%. Our experimental results of classification and segmentation achieve 100% accuracy and 0.882 dice coefficient of tumor region respectively, which are better than the results of a single classification network and segmentation network.

Index Term - Convolution neural network, multi-task, semantic segmentation, medical image processing

I. INTRODUCTION

As one of the top 10 prevail cancers in the world, renal tumor receives wide attentions from medical experts and concerned people. Renal tumor is not sensitive to radiotherapy and LPN is the state-of-the-art therapy. However, limited to the situations of benign tumors and low-grade ones [1], LPN must be based on accurate classification and precise localization of tumor before operation.

Angiomyolipoma and clear renal cell carcinoma are the primary subtypes of benign and malignant tumor separately. Quite resembling other subtypes, angiomyolipoma with minimal fat (AMF) is intractable to be differentiated. To distinguish this kind of tumor, some works [2], [3], [4] have been proposed. Besides, some researches [5], [6] also focused on classifying other subtypes. Most of those methods incorporate a 2-step framework. In the framework, some manually selected operators are used to get features of tumor firstly. Then, SVM is employed as a classifier to get classification results. Although this framework can handle

some tasks, we can find that it dose not work effectively and perform well since it is trained based on small dataset and manually selected features.

In general, the majority of researches on kidney images are concentrated on segmentation. Khalifa et al. [7] introduced geometric deformable models and clustering for 3D kidney segmentation in CT images. Marsousi et al. [8] used PKSM to locate the best matching position to detect kidney and fitted it by a global deformation. Although existing works have got good performance on the segmentation of kidney region,various intensity distribution, erratic location, irregular shape, etc bring difficulties to the segmentation of tumor.

Deep learning has achieved some success in medical issues [9]. Our previous work [10] leveraged a 3D fully convolutional neural network to the segmentation of renal tumor. Although that work got good results, more subtypes of tumor and multi-phasic CT images can be explored. Zhou et al. [11] put forward that a convolutional neural network with larger receptive field would be better on learning features from objects of different scales in one image. The theory of enlarging receptive field appears well adapted to problems with which the segmentation of renal tumor faces. Aimed at expanding receptive field, some works were proposed, such as dilated convolution [12], spatial pyramid pooling [13] and PSPNet [14].

In this paper, we try to utilize features from a segmentation network based on PSPNet as the feature extractor of a classification network and combine classification and segmentation network together. The novel points of our approach are: (1) Compared to single networks, the combination of segmentation and classification improves both results and is more suitable to our task. Classification results have a positive impact on segmentation results and segmentation results can give the classification network information of RoI in return. (2) In the segmentation network, we adopt a 2-step segmentation strategy which gives an attention to the network. Such attention leads to a boost in accuracy.

II. METHODOLOGY

A. Network structure

The original images of renal tumor are 3D CT scans in which one image owns around 300 slices. A 3D network as we did [10] based on 3D PSPNet is not operant to our multi-tasks because of the huge demand of graphics memory. Thus, in this work, 3D images are converted into 2D and the neural network is 2D CNN too.
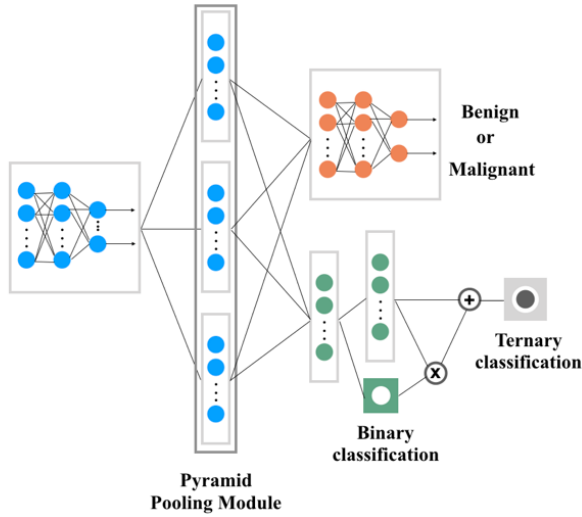
Fig. 1.   The structure of SCNet

Fig. 2.   The structure of pyramid pooling module

The main structure of SCNet is shown in the Fig. 1. It is consisted of three networks, basic convolutional neural network (BNet), segmentation network (SegNet) and classification network (ClsNet). The blue part is the backbone PSPNet which comprises ResNet50 and pyramid parsing module (PPM) showed in section II-B. Following PPM, there are two modules, ClsNet in orange and SegNet in green. In ClsNet, several residual blocks are adopted to fuse features to classify tumor types. The structure of SegNet is detailed in section II-C.

Considering that the size of CT scans, 150×150, is smaller than the size of input of the original ResNet50, we decrease the number of pooling layer and avoid to use convolutional layers with stride 2. The decline of pooling layer will decrease receptive field. Thus, dilated convolution is also applied to the network to magnify receptive field.

B. Pyramid parsing module

As mentioned in section I, we can expand receptive field of the network to extract features at different scales. Stemming from PSPNet, PPM has different down-sampling sizes which are benefit to learning multi-scale features.

This module has several branches which down-sample the input features in different sizes. Through down-sampling, convolution and up-sampling, all outputs of branches have same size of feature map. After that, those outputs are merged into a new tensor. Fig. 2 shows the structure of it. In the proposed network, down-sampling ratios are set as 1/2, 1/4 and 1/8 respectively and the module has a direct short-cut from input.

C. 2-step segmentation strategy

For segmentation, a two-step segmentation strategy is proposed. After PPM, as the Fig. 3 shows, firstly, a 2-class classifier and a convolutional block are calculated separately. The 2-class classifier distinguishes kidney and tumor from background. The channel of the convolutional
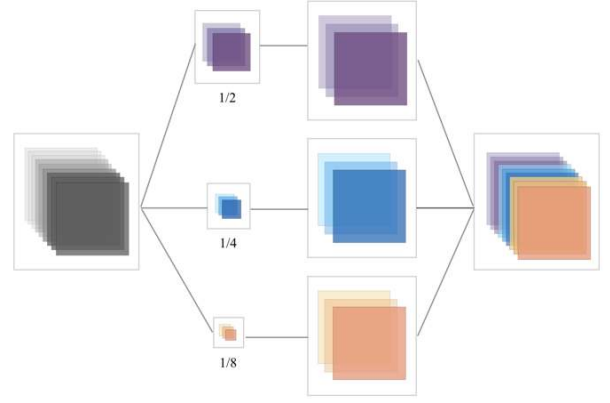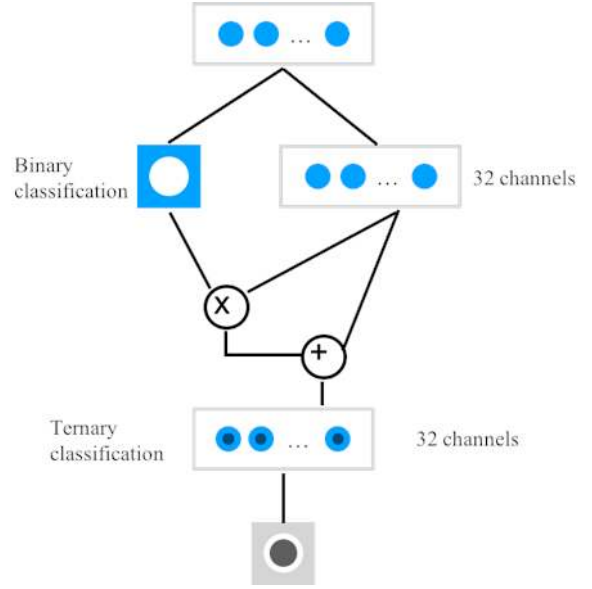
Fig. 3.   The 2-step segmentation strategy

block is 32, which is aimed at reducing computation. In this way, a probability map of kidney and tumor region and 32 feature maps are obtained. Then 32 feature maps are multiplied by the probability map. This multiplicative operation can enlarge the value of RoI and widen the gap between background and RoI. Finally, 32 enhanced feature maps are classified into 3 classes, namely, background, kidney and tumor. We designate this two processes as binary classification and ternary classification separately. By this strategy, binary classification results offer the network an attention and give the ternary classification enhanced results. The process is summarized in equation (1)

$$Y = \Phi_2(\Phi_1(X) \times F(X) + F(X)) \qquad (1)$$

where Φ_2 and Φ_1 are the 3-class classifier and 2-class classifier respectively. F is the 32-channel convolution layer. X and Y are input and output.

D. Loss function

Totally, the cost of SCNet is the sum of costs from SegNet and ClsNet as equation (2) shows. For the cost of SegNet, it unites cost values of binary and ternary classification. As for optimization equation, cross-entropy loss function explained in equation (4) is used for all items

$$L_{total} = L_{SegNet} + L_{ClsNet} \tag{2}$$

$$L_{SegNet} = L_{binary} + L_{ternary} \tag{3}$$

$$L = -\sum_{x} p(x) \log q(x) \tag{4}$$

III. EXPERIMENTAL RESULTS

A. Experimental dataset and implementation details

The effectiveness of the proposed method is evaluated on 3D CT images supported by radiology department of Jiangsu Province Hospital. The dataset is comprised of four subtypes of renal tumor, i.e. benign AMF and malignant renal cell carcinomas (RCCs) which are clear RCCs, papillary and chromophobe. The benign subtype only accounts for 23%. The dataset is consisted of image pairs including CTA (CT angiography) and CTU (CT urography) images.

There are 30 3D images of benign tumor and 101 3D images of malignant RCCs in the dataset. The ground truth of those medical images were drawn by experts. 3D images were transferred to 2D images. Rotation, scaling, cutting, translation and so on were operated to images to augment data.

Because the dataset is unbalanced, we apply different times of augmentation to malignant and benign tumors to get a balanced one. Finally, the data is augmented 16 times larger than it before.

After transferring 3D images into 2D, how to identify malignant and benign type of slices which have no or a speck of tumor tissue is a significant concern. We give those special 2D images a different label normal. The dataset has 3 types of images, benign, malignant and normal.

Our work was implemented on pytorch (http://pytorch.org/) and performed on a workstation with the CPU of i7-5930K, the RAM of 128GB and four graphic cards of TITAN X of 12GB memory. The learning rate started at 1e-3 and the network converged after 6 epochs.

B. Evaluation results

Algorithms of segmentation and classification are compared separately. Results of all methods as well as SCNet are summarized in Table I. For classification of benign and malignant, a 10-fold cross validation is carried out to evaluate all classification networks. For segmentation, 1/4 of dataset is kept for testing and the rest is used as training data. For models of classification, we adopt 10-fold cross validation. For models of segmentation, tests are based on test set.

TABLE I

RECALL RATES AND DICE COEFFICIENT OF CLASSIFICATION AND SEGMENTATION RESULTS OF DIFFERENT ALGORITHMS AND MODELS.

Algorithms	Classification		Segmentation	
	Malignant	Benign	Kidney	Tumor
SVM	0.950	0.851	-	-
ResNet50	0.851	0.90	-	-
U-NET	-	-	0.907	0.911
PSPNet	-	-	0.799	0.848
SCNet + ternary loss	1.0	1.0	0.945	0.854
SCNet + 2-step loss	**1.0**	**1.0**	**0.944**	**0.882**

For classification, there are many effective networks, such as VGG16 [15], ResNet and so on. We use ResNet50 as a contrast network. Considering that traditional machine learning played an important role in medical image processing, SVM is also considered. Some first order statistics, features based on shape and size and textural features are extracted as the feature group. The SDK LIBSVM [16] was utilized to realize a SVM classifier.

In test, all slices of a patient will be sent into SCNet model to get classification results, the same as other networks do. Then, the benign or malignant class of the whole 3D image is determined only by the classification results of slices with tumor judged by the network. In our case, the tumor type of one patient is determined by majority vote.

As table I shows, SCNet achieves the classification accuracy 100% in both benign and malignant types, which is higher than SVM up to 5.0% and 13.3% separately, meanwhile, 14.9% and 10.0% higher than ResNet50. It's noteworthy that SVM got better classification result in malignant tumors. Trained on unbalanced dataset, CNNs faced a overfitting dilemma. SCNet got higher precision because the combined loss can prevent the network from overfitting.

For segmentation, SCNet, PSPNet and U-Net were compared. Input layers and output layers of those available networks were adjusted to adapt to our data. In addition, the baseline of PSPNet is ResNet50 while SCNet is the same.

As the table shows, SCNet has achieved dice coefficient 0.944 and 0.882 separately. Compared to PSPNet, the segmentation result of tumor was improved 3.4%. This result proves that classification module is beneficial to segmentation network.

Moreover, we did a comparative analysis about implementations of ternary loss and 2-step loss individually, which are discussed in section II-C. Obviously, 2-step loss got a more satisfactory result that the dice coefficient of tumor was improved by 2.8% against its counterpart.

C. Discussion

From results, it can be observed that SCNet performs the best outcomes in both segmentation and classification. Fig. 4 shows the segmentation results of different models. Compared to PSPNet and U-Net, results of SCNet have smoother edges and more precise segmentation.

For further statistics, a histogram is drew which displays relations between tumor size and segmentation accuracy.
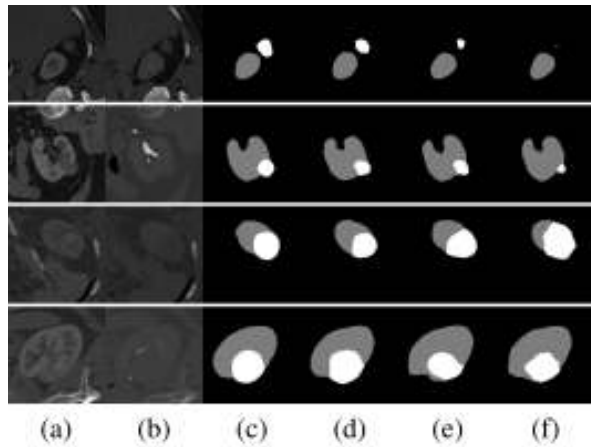
Fig. 4. Segmentation results of different models. Here (a) CTA (b) CTU (c) ground truth (d) SCNet + 2-step loss (e) PSPNet (f) U-Net.
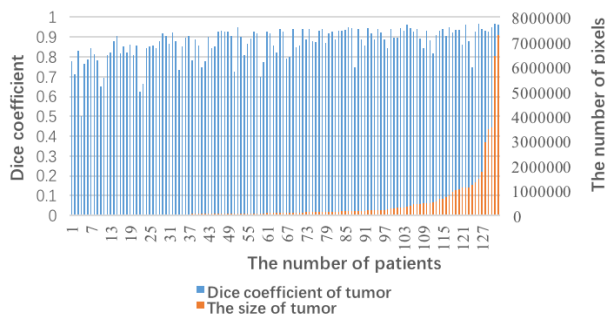

Fig. 5. The histogram of segmentation results and tumor size. Here, the number of pixels of tumor region is seemed as the tumor size.

Fig. 6. The histogram of classification results and tumor size. Here, the accuracy of one patient is calculated based on all slices.

In our case, the number of pixels of tumor region in the whole 3D image for one patient is seemed as the tumor size. From the Fig. 5, a reasonable result can be obtained, that the smaller tumors are, the worse segmentation results are. For example, only five patients, whose dice coefficients of segmentation results are lower than 0.7, are characterized by small tumor size (index 4, 10, 11, 22 and 23). As we can see from the figure, when the size of tumor increases, dice coefficient of segmentation results is improved as well.

The average of classification results on 2D images are 97.3% as Fig. 6 shows, which means that SCNet performs well in classification too.

IV. CONCLUSION

In total, SCNet is an effective multi-task network and results are improved remarkably. It is also a flexible network. For example, DenseNet, VGG16 or some other networks can be used as a backbone too. The 2-step segmentation strategy can also be transferred to other networks to improve results.

REFERENCES

[1] Georges-Pascal Haber and Inderbir S Gill, "Laparoscopic partial nephrectomy: contemporary technique and outcomes," *european urology*, vol. 49, no. 4, pp. 660–665, 2006.
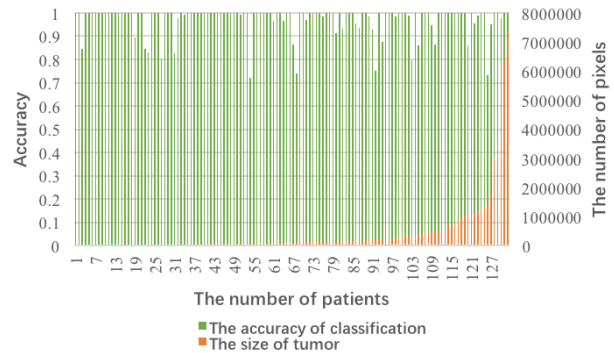
[2] Claus Simpfendorfer, Brian R Herts, Gaspar A Motta-Ramirez, Daniel S Lockwood, Ming Zhou, Michael Leiber, and Erick M Remer, "Angiomyolipoma with minimal fat on MDCT: can counts of negative-attenuation pixels aid diagnosis?," *AJR. American journal of roentgenology*, vol. 192, pp. 438–443, Feb. 2009.

[3] Humaira S Chaudhry, Matthew S Davenport, Christopher M Nieman, Lisa M Ho, and Amy M Neville, "Histogram analysis of small solid renal masses: differentiating minimal fat angiomyolipoma from renal cell carcinoma," *American Journal of Roentgenology*, vol. 198, no. 2, pp. 377–383, 2012.

[4] Ching-Wei Yang, Shu-Huei Shen, Yen-Hwa Chang, Hsiao-Jen Chung, Jia-Hwia Wang, Alex TL Lin, and Kuang-Kuo Chen, "Are there useful CT features to differentiate renal cell carcinoma from lipid-poor renal angiomyolipoma?," *American Journal of Roentgenology*, vol. 201, no. 5, pp. 1017–1028, 2013.

[5] Taryn Hodgdon, Matthew DF McInnes, Nicola Schieda, Trevor A Flood, Leslie Lamb, and Rebecca E Thornhill, "Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images?," *Radiology*, vol. 276, no. 3, pp. 787–796, 2015.

[6] Jonathan R Young, Daniel Margolis, Steven Sauk, Allan J Pantuck, James Sayre, and Steven S Raman, "Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT," *Radiology*, vol. 267, no. 2, pp. 444–453, 2013.

[7] Fahmi Khalifa, Ahmed Soliman, Ali Takieldeen, Mohamed Shehata, Mahmoud Mostapha, Ahmed Shaffie, Rosemary Ouseph, Adel Elmaghraby, and Ayman El-Baz, "Kidney segmentation from CT images using a 3D NMF-guided active contour model," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 432–435.

[8] Mahdi Marsousi, Konstantinos N Plataniotis, and Stergios Stergiopoulos, "Shape-based kidney detection and segmentation in three-dimensional abdominal ultrasound images," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 2890–2894.

[9] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps*, pp. 323–350. Springer, 2018.

[10] Guanyu Yang, Guoqing Li, Tan Pan, Youyong Kong, Jiasong Wu, Huazhong Shu, Limin Luo, Jean-Louis Dillenseger, Jean-Louis Coatrieux, Lijun Tang, et al., "Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3790–3795.

[11] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba, "Object detectors emerge in deep scene cnns," *arXiv preprint arXiv:1412.6856*, 2014.

[12] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene

parsing network," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 6230–6239.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.