Pyramid Spatial Pooling Convolutional Network for whole liver segmentation

Jessica C. Delmoral[1]
jessica.delmoral@fe.up.pt
Diogo B. Faria[2]
dborgesfaria@gmail.com
Durval C. Costa[3]
durval.c.costa@gmail.com
João Manuel R. S. Tavares[1]
tavares@fe.up.pt

[1] Instituto de Ciência e Inovacão em Engenharia Mecânica e Industrial, Porto, Portugal
[2] School Of Health Sciences - University of Aveiro
[3] Champalimaud Centre for the Unknown, Fundação Champalimaud, Portugal

Abstract

Segmentation of the liver in Computer Tomography (CT) images allows the extraction of three-dimensional (3D) structure of the liver structure. The adequate receptive field for the segmentation of such a big organ in CT images, from the remaining neighboring organs was very successfully improved by the use of the state-of-the-art Convolutional Neural Networks (CNN) algorithms, however, certain issue still arise and are highly dependent of pre- or post- processing methods to refine the final segmentations. Here, an Encoder-Decoder Dilated Poling Convolutional Network (EDDP) is proposed, composed of an Encoder, a Dilation and a Decoder modules. The introduction of a dilation module has produced allowed the concatenation of feature maps with a richer contextual information. The hierarchical learning process of such feature maps, allows the decoder module of the model to have an improved capacity to analyze more internal pixel areas of the liver, with additional contextual information, given by different dilation convolutional layers. Experiments on the MICCAI Lits challenge dataset are described achieving segmentations with a mean Dice coefficient of 95.7%, using a total number 30 CT test volumes.

1 Introduction

Automatic segmentation of different medically relevant liver tissues is continuously an active research topic in medical image analysis. Such segmentations can provide doctors with meaningful and reliable quantitative information of the structure of the liver, which subsequently enable the identification of abnormalities. Knowledge of the liver structure becomes particularly relevant in individuals diagnosed with liver cancer. In this clinical scenario, physicians need to study the full liver physiology and make an informed decision on the treatment course. Liver cancer treatment may include chemo- or radio- therapy, hepatectomy (liver resection) or in very specific cases transplantation. Liver cancer has an alarming prevalence in a global scale and is the second most lethal cancer worldwide, accountable for more than 788,000 deaths in 2015 [6]. Computed Tomography (CT) is one of the most common modalities used for detection, diagnosis and follow-up of liver cancer [4]. Liver cancer is characterized by the development of abnormal cell accumulations, commonly referred as lesions that will appear represented differently within the anatomy of the liver, in structural images such as Computed Tomography (CT). Thus, in the clinical setting the image-aided diagnosis requires an accurate segmentation of the whole liver anatomy in CT images. In this paper, we present a new method for training a global CNN for liver segmentation in CT scans address the issues above by developing a fully automatic liver segmentation model which efficiently combines the FCN with residual blocks and dilated convolutions.

2 Methods

2.1 Dataset

Public datasets are commonly used for assessing liver cancer in CT tissue segmentation algorithms as they provide ground truth labels. The model studied was trained and tested on data from the 2017 LITS MICCAI challenge dataset. We use a total of 130 CT volumes. Each image volume was characterized by a 512x512 image resolution and varying number of slices, ranging from 91 to 844. Having in mind the segmentation problem in this setting is regarded as a pixel-wise classification, the classification targets are comprised by two different classes: liver and background. To prevent extra errors induced by class data imbalances, the models proposed here were trained only on central slices depicting the liver area, comprising a total of 90 slices from each exam. CT images acquisition outputs a quantification of X-rays by tissues at a pixel wise level, which is outputted according to the known scale of Hounsfield units (HU), proportional to the degree of tissue attenuation suffered. Although different HU intervals characterize different organs, these values often overlap, difficulting the intuitive discrimination of the present tissues. To eliminate the noise effect of other HU value intervals, a technique named CT windowing is often applied. Thus, all CT slices were thresholded with a window range of [–200, 200] recommended for a liver to remove the irrelevant tissue intensities.

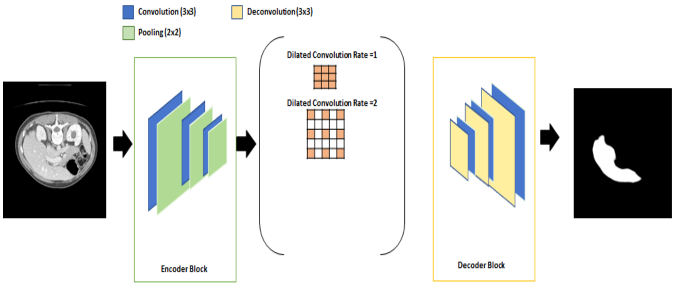

Figure 1: Neural network architecture of the Encoder-Decoder Dilated Poling FCN (EDDPFCN).

2.2 Feature learning of the proposed CNN model

Input images and the corresponding liver segmentation masks provided by human experts were used to train the network. Examples of ground truth masks are latter presented in the Results section. To learn the whole liver supervised features an FCN model was trained. Such model was formulated taking into account the several sizes of receptive fields that can allow the network to learn the most discriminative feature maps. Such methods require also the adequate the kernelized image context to correctly identify the liver voxels. The size of receptive field roughly indicates the amount of context information that is used in each feature map.

The proposed architecture is based on the state-of-the-art segmentation architecture called "Atrous Spatial Pyramid Pooling" (ASPP). The network works with 2D slice-wise axial images and is composed of (a) three initial Residual Convolutional-Pooling blocks, followed by (b) five parallel layers of dilated convolutions with rate r = 1,2,4,6,8,16,32 which were concatenated and forwarded to (c) three deconvolutional blocks. The network outputs are fine-tuned with a Sigmoid layer using the given labels. Given a 1D signal x[i], the y[i] output of a dilated convolution with the dilation rate r and a filter w[s] with size S is formulated as

$$y[i] = \sum_{s=0}^{S} x[i + r \cdot s]w[s].$$

Such type of convolutional kernel is also rotation invariant. The dilations can be mentally conceptualized as the introduction of discrete intervals of pixel that are used for the convolution kernel, that are dictated by the dilation rate r. Figure 1 illustrates the pipeline of the proposed training process.

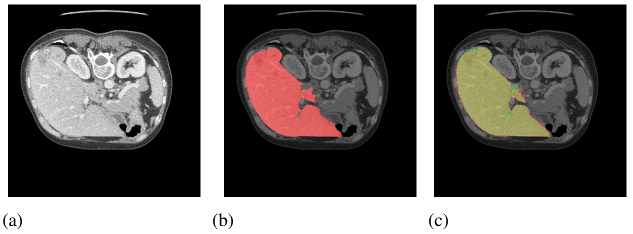

Figure 2: Neural network architecture of the Dilated Spatial Pooling FCN (DSPFCN). (a) represents the original CT, (b) the ground truth mask; and (c) the overlay of the obtained mask and the ground truth.

2.3 Model training and parameter fine-tunning

The model parameters are learnt from training data by minimizing a loss function, via end-to-end training. Given the chosen dataset division a total of 8100 image slices, from 100 subjects were used as training samples. For training we use the Adaptive Moment Estimation (Adam) algorithm [5] for model optimization during 70 epochs with learning rate reduced by a factor of 0.9 of the original value after 1/3 rd and 2/3 rd of the training finished. The network weight initialization with and without He norm initialization were explored. The dataset was augmented using rotation and horizontal flipping to increase generalizability of the model. The hyper-parameters were tuned so as to give best performance on validation set. Training took 8 hours on one Nvidia Titan XP GPU. During training a set of 900 image slices were used as a validation dataset to monitor the models performance evolution during training. Having previous knowledge unbalanced class distribution between the liver and background, we implement a loss function that attempts to overcome this limitation. To this end, the Dice Similarity Coefficient Loss was chosen for objective function minimization during model training. Such loss function has been extensively validated in the literature for Convolutional Neural Networks training, due its insensitivity to class imbalancing. The model inputs each image slice, of size 512x512 individually using only contextual information in the orthogonal direction. The model outputs the pixel-wise classification into each of the three classes.

3 Results

In this section, we have discussed the results of the pixel-wise classification of the images. We trained the model on 3D CT data for liver segmentation. Table 1 shows the comparative test results of the proposed model and the top performing methods in the literature. A total 2700 image slices, corresponding to a total of 30 3D CT scans, not previously used for model training, were used to test the performance of the proposed model. The qualitative results of the segmentation results can be evaluated through Figure 1. In the two example results, the complex and heterogeneous structure of the liver were detected in the shown images. Overall, the model predictions were accurate in the classification of true positives. However, from the analysis of the entire dataset, the fuzyness of the liver boundaries in some scans leaked to the neighboring tissues, depicted in similar intensities. This is observable in Figure 2, in the example (c). To quantitatively evaluate the classification performance, we report the segmentation quality results of two metrics, proposed previously in the literature namely, the Dice (DSC) and Jaccard (JC) coefficients.

4 Discussion

In this work, we devise a simple, but efficient and end-to-end method that achieves state-of-the-art results in both quantitative metrics when compared to the four top performing methods of the literature, as detailed in Table 1. To the best of our knowledge, no previous method taking advantage of the positive performance aspects of dilated convolutions was previously proposed for the task of liver segmentation in abdominal CT images. In medical imaging, the most traditional architecture for segmentation is the well-known U-Net, which is characterized by two distinct sequential blocks of encoder and decoder or contracting and expanding convolutional segments that basically aggregate semantic informations. The simplicity of the proposed method when compared to some of the most traditional methods used such as the U-Net [3], provided 1) better performing results, but also 2) a parameter reduction that is achieved by the efficiency of the inclusion of the dilated convolutions.

Table 1: Liver segmentation performance using different algorithms

Method	Dice(%)	Jaccard (%)
[2]	95.90	92.19
[1]	89	
[3]	94.3	-
[7]	96.3	-
Ours	95.7	91.3

5 Conclusions

In the present work, a novel CNN architecture for whole liver segmentation in CT images is proposed. The segmentation of a big organ such as the liver, would in many previous architectures be penalized by an inadequacy of the receptive field used for feature learning in previously proposed architectures. The key concatenation of dilation convolutions has allowed accurate segmentations of the final liver boundaries, with minimal fuzziness. No hole filling post-processing was needed with the proposed architecture. In future works, the proposed architecture potential to segment other liver tissues, such as lesions and vascular structure will be explored. Moreover, advanced techniques of data augmentation using adversarial networks, could further improve the resulting segmentations obtained in the present study.

Acknowledgments.

The authors gratefully acknowledge the funding from Project NORTE-01-0145-FEDER- 000022 - SciTech - Science and Technology for Competitive and Sustainable Industries, cofinanced by "Programa Operacional Regional do Norte” (NORTE2020), through "Fundo Europeu de Desenvolvimento Regiona” (FEDER). The authors also kindly thank Nvidia, for the contribution with one Nvidia Titan XP GPU, that was used in this work.

References

[1] A Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan. Fully convolutional network for liver segmentation and lesions detection. In G et. al Carneiro, editor, *Deep Learning and Data Labeling for Medical Applications*, pages 77–85, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46976-8.

[2] Lei Bi, Jinman Kim, Ashnil Kumar, and Dagan Feng. Automatic liver lesion detection using cascaded deep residual networks. 2017. URL `http://arxiv.org/abs/1704.02703`.

[3] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, et al. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*, 2017.

[4] Lucy E. Hann, Corinne B. Winston, Karen T. Brown, and Timothy Akhurst. Diagnostic imaging approaches and relationship to hepatobiliary cancer staging and therapy. *Seminars in Surgical Oncology*, 19(2):94–115, 2000. ISSN 1098-2388.

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[6] World Health Organization. Fact sheet: cancer, 2015. URL `http://www.who.int/mediacentre/factsheets/fs297/en/`.

[7] Yading Yuan. Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation. 2017. URL `http://arxiv.org/abs/1704.02703`.