Segmentation of Kidney and Tumor using Auxiliary Information

Muhammad Usman Akbar[1,2], Vittorio Murino[1,3], and Diego Sona[1,4]

[1] Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genova, Italy,
[2] Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture, Università degli Studi di Genova, Italy
[3] Department of Computer Science, Università di Verona,
[4] Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy,
`muhammad.akbar@iit.it`

Abstract. Automatic segmentation of organs and tumors is a prerequisite of many clinical application in radiology. The anatomical variability of organs in the abdomen and especially of tumors makes it difficult for many methods to obtain good segmentations. in this report we present a cascade of two convolutional neural networks allowing to segment an organ followed by the segmentation of a tumor. The advantage of the proposed pipeline is that the preliminary organ segmentation, which is a simpler task, helps the further segmentation of the tumor. The proposed system was evaluated using the KiTS19 challange dataset.

Keywords: Deep learning · ensemble learning · convolutional neural networks · medical imaging · segmentation · kidney · tumor

1   Introduction

Accurate segmentation of tumors is a fundamental step in many automated clinical applications. Manual segmentation is still the standard approach, which however is time consuming, subjective and prone to errors due to both the inherent subjectivity of the expert and the fluctuations in human efficiency. With the advancement in technology many semi-automated systems have been introduced, which are very helpful but still they need the possibly biased human interaction.

To overcome this problem various automated techniques were introduced. Recently, deep convolutional neural networks (CNNs) have proven to be very effective in many tasks including segmentation, outperforming many state-of-the-art traditional approaches. In general, all deep learning methods work with two different approaches, either they process the full image (in 3D or slice by slice) [1] or they work with a patch based approach where multiple small patches (in 2D or 3D) are processed separately and results are concatenated to reconstruct the segmentation at the original size [3].

Our aim is to create a pipeline of CNNs working on the full image by slices, which uses two models in order to simplify the complex segmentation of tumors using the outcome of the simpler organ segmentation.

The paper is organized as follows. The proposed system is first explained in section 2, together with a description of the used dataset and the experimental setup. Finally, some conclusion will be drawn in section 4.

2   Data and Methods

2.1   Kidney Tumor Segmentation Dataset

We used the dataset provided in the challenge KITS'19 [5]. Totally there are 300 subjects, which are divided into a training set of 210 subjects and a test set of 90 subjects. The size of individual slices in the volume is 512x512 pixels and the number of slices of the dataset varies from 31 to 1059. The 210 samples in the training dataset are provided with a ground truth labeling distinguishing between background, kidney and kidney tumor.

2.2   Proposed System

The proposed system is built using two models working separately on the two classes (kidney and tumor). In an ensemble learning perspective, instead of using traditional approaches to combine the outcome of multiple models, we use the two models in a cascade, using the prediction calculated from first model as auxiliary information to the input of the second model. The first model is used to compute the prediction of the kidney, which is then added as auxiliary input channel to the original image. Hence, the two channel image is used in the second model to produce a prediction of tumor segmentation. The principle is that the tumor segmentation is a more complex task as the involved structures are small. Hence the knowledge where the kidney is helps the system to understand where to search for the tumor. Once both predictions are calculated they are merged together to generate the final prediction. The flow chart of the proposed system can be seen in Figure 1. Both models are based on Fc-DenseNet56 [2], The model we are using is FC-DenseNet56 for computing both predictions and in this model feature maps were extracted in dense block of transition down layer. Pre-activation layer was used where ReLU, convolution, max pooling and Batch normalization were performed on the input slice of 512x512. Up-sampling was performed in transition up layer where input was up sampled and concatenated with the skip connections and final prediction was computed using soft-max layer.

2.3   Experimental Setup

The model used to compute the segmentation of kidney was trained in 10 epoch, while the model used to compute the segmentation of tumor was trained in

---
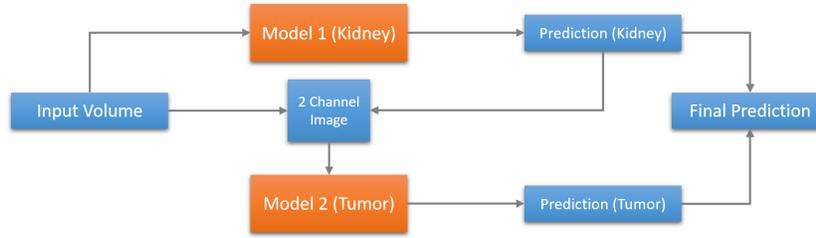
Fig. 1. The processing pipeline composed by two CNNs. First Model computes the prediction for kidney segmentation, which is then used as auxiliary information to the original image to predict the tumor segmentation. Finally, both the predictions of kidney and tumor are joined together to get the final semantic segmentation.

15 epoch. RMSPropOptimizer was used for both models with a learning rate of 0.0001. The proposed system was implemented on NVidia GTX 1080. The training was done from scratch, 40 subjects from 210 subjects were used for validation and for evaluation purpose, we used Dice score to measure the intersection between resulting segmentation and ground truth. The predictions computed for the test dataset have been submitted to the challenge website, which then provided the performance. The main steps of the proposed methodology are mentioned below.

- **Pre-processing:** individual ground truth volumes for kidney and tumor were created from the original image. No other pre-processing was applied to original images.
- **Setting training for kidney:** all slices from all original volumes and corresponding ground truth labeling of kidney were used to generate the training and validation dataset for the first model.
- **Setting training for tumor:** two channel images were generated concatenating the original image slices and the corresponding prediction for kidney segmentation (binary images). These two channel images were then used to train the second model, again 170 subjects were used for training and the remaining 40 subjects were used for validation.
- **Final prediction:** once the prediction from the first model (kidney) and second model (tumor) were computed they were merged to obtain the subject-wise volumes containing the semantic segmentation of kidney and tumor.

3   Results

Dice score on the validation data was used to stop the training of the models. The mean Dice score calculated for kidney and tumor validation subjects was 0.924 and 0.484 respectively. The trained models were used to generate the final prediction of the test subjects provided by the competition website. Dice score

for the test subjects were provided by the competition management and the approximate dice for kidney and tumor was 0.932 and 0.414 respectively. The average kidney-tumor approximation was calculated to be 0.673.

4   Conclusion

In this paper, we proposed an architecture built upon the combination of two models, which are connected in order to let the first model help the second model to obtain better prediction. This is obtained providing the kidney segmentation from the first model as auxiliary information to the second model for tumor segmentation. The auxiliary information resulted to be useful improving the system performance.

References

1. Aslani, S., Dayan, M., Storelli, L., Filippi, M., Murino, V., Rocca, M.A., Sona, D.: Multi-branch convolutional neural network for multiple sclerosis lesion segmentation. NeuroImage **196**, 1–15 (2019)
2. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 11–19 (2017)
3. Wang, Z., Bhatia, K.K., Glocker, B., Marvao, A., Dawes, T., Misawa, K., Mori, K., Rueckert, D.: Geodesic patch-based segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 666–673. Springer (2014)