A two-stage network for cardiac segmentation with adapted data augmentation

Georgios Simantiris simge@csd.uoc.gr and Georgios Tziritas tziritas@csd.uoc.gr

Computer Science Department, University of Crete

Abstract

We present a method for semantic segmentation in cardiac MR images using a two-stage convolutional neural network. We work in full resolution throughout the network’s layers utilizing dilated convolutions for preserving localization accuracy. Specially designed data augmentation is invoked. We apply the methodology for left ventricle cavity and myocardium segmentation with promising results.

Keywords: Semantic segmentation, Deep Learning, Dilated convolutions, Left Ventricle.

1. Introduction

In recent years Deep Convolutional Neural Networks (CNNs) have achieved state-of-the-art results in biomedical image segmentation tasks (Ronneberger et al., 2015). Approaches with CNNs for the segmentation of the Right and Left Ventricles, as well as the myocardium (Wolterink et al., 2018; Tran, 2016; Isensee et al., 2018), and for subsequent LV quantification in cooperation with Recurrent Neural Networks (RNNs) (Xue et al., 2018), resulted in excellent performance.

In this paper we address the task of semantic segmentation of the Left Ventricle (myocardium and LV cavity area) designing a new two-stage CNN architecture with dilated convolutions as elaborated in Section 2. We train the network end-to-end utilizing a novel data augmentation technique presented in Section 3. First experiments and preliminary results are discussed in Section 4.

2. Methodology

We propose a two-stage CNN, which is designed to gradually refine its predictions. Essentially, the first stage aims to estimate the joint data/label distribution, while the second improves the prediction by better learning the high order context of the label field. We choose to work in full resolution at all network layers in order to accurately localize the region boundaries, as sub-sampling or pooling might cause high-frequency details losing. This is very important for various applications, because image segmentation could be used subsequently for deriving accurate quantitative measurements for other decisions.

The step-wise approach is assisted by incorporating the loss function into each step, so that respective inference contributes with a proper weight to the final prediction (Yu et al., 2017). The modified Dice score loss function is used for preventing that classes with larger regions dominate over classes with smaller regions by counteracting with weight factors, thus
allowing the network to learn to accurately segment smaller regions as well (Crum et al., 2006; Sudre et al., 2017). Also, anatomical constraints are imposed in the sense that the left ventricle cavity must be surrounded by the myocardium and therefore predictions which violate this constraint are also punished resulting in a higher loss.

![Two-stage network diagram](image)

**Figure 1: Two-stages network.**

The proposed network’s architecture is shown in Figure 1. The two main stages are separated with a dashed red vertical line, where at this point only the estimations of the likelihood probabilities are fed to the latter stage in opposite to the former stage where all the feature maps are used at each step. In the first stage we start with bigger kernel sizes, gradually reducing them while increasing the dilation factor. In the second stage smaller kernels start, but are in the next step increased, in opposite to the dilation factor which is reduced. The total number of parameters of the network is 252447.

3. Data augmentation

It is common in all Deep Learning approaches, where the training samples are limited to a small number, to invoke extensive data augmentation to prevent overfitting. That is also necessary in our case and we opted for an adaptive approach. In order for the network to learn how to accurately distinguish the left ventricle (LV) cavity from the myocardium, the intensities of these two areas are transformed to produce augmented images. The probability density functions of the two areas, myocardium and LV cavity, are estimated for bringing the intensities closer to each other. The contrast reduction is adjusted by the Bhattacharyya distance between the two distributions. By also adding noise, finally harder cases of training samples are obtained.

4. Experiments and Results

The network was trained using the Cardiac-DIC dataset (LVQ, 2018) with 2900 images and their corresponding ground truths obtained from SAX MR sequences of 145 subjects, where
for each subject 20 frames were included. Since this dataset was originally created for LV quantification and the accompanied test set serves the same purpose, we randomly selected 29 subjects, thus 580 images, to form a Test dataset.

The remaining 2320 images of 116 subjects were augmented as described above, in a preprocessing procedure to form a training set of 23200 images in total. An Adam optimizer was used to update network weights. Accuracy, in the sense of how often predictions have maximum in the same class as the ground truth, served as the metric. Larger batch sizes performed better, therefore training was held in batches of 64 images for 20 epochs only, since it was enough for the network to learn. Overfitting was furthermore tackled by also using Batch Normalization and L2 regularization.

7-fold cross-validation training showed consistent network performance. The mean training loss was 0.0033 while the mean validation loss stayed close enough at 0.0044. After training the weights of the best performing fold were used to obtain predictions for the test set, and the mean Dice similarity coefficient measured over all the 580 images was 0.8907 for the myocardium and 0.9553 for the LV cavity. Predictions can be seen in Figure 2, where outputs of the steps are also depicted showing that the network starts from a rough probability estimation and gradually refines this a posteriori probability yielding in a promising segmentation.

![Figure 2: Original images and gradual network outputs. The final image depicts the ground truth boundaries and the network predictions. LV cavity: red/white, myocardium: green/gray.](image)

### 5. Conclusion

We have introduced a two-stage convolutional neural network for inferring the likelihoods and adjust them in order to perform accurate semantic segmentation in medical images. The network, while still at its early development, seems robust and cohere. We plan to further develop, adapt and test it with other datasets, as well as different number of classes to be segmented.
References


