Using Deep Learning to Estimate Systolic and Diastolic volumes from MRI-images

Jonas Degrave JONAS.DEGRAVE@UGENT.BE
Jeroen Burms JEROEN.BURMS@UGENT.BE
Iryna Korshunova IRYNA.KORSHUNOVA@UGENT.BE
Joni Dambre JONI.DAMBRE@UGENT.BE
Ghent University, Technologiepark 15, 9052 Zwijnaarde, Belgium

Keywords: Deep Learning, MRI, Diagnostics, Ejection Fraction

Abstract

Deep learning with convolutional neural networks has become a widely used tool for computer vision tasks. In this paper, we focus on the problem of automatically annotating the volumes of the left ventricle in the heart, based on DICOM MRI-images. We discuss the solution, which was the second best in an international data mining competition. We show that it is possible to achieve near-human performance using a deep learning approach when using task-specific model architectures.

1. Introduction

The goal of this year's Data Science Bowl[1] was to estimate minimum (end-systolic) and maximum (end-diastolic) volumes of the left ventricle from a set of MRI-images taken over one heartbeat. These volumes are used by practitioners to compute an ejection fraction: the fraction of outbound blood pumped from the heart with each heartbeat. This measurement can predict a wide range of cardiac problems. For a skilled cardiologist, analysis of MRI scans can take up to twenty minutes. Therefore, automating this process would be useful to decrease the cost of medical care.

2. Dataset

The dataset consisted of over a thousand patients, only 500 of which where provided as initial training set. For each patient, we were given a number of 30-frame MRI videos in the DICOM format, such as shown in Figure 1, showing the heart during a single cardiac cycle. These videos were taken in different planes including the multiple short-axis views, a 2-chamber view, and a 4-chamber view. The SAX views, whose planes are perpendicular to the long axis of the left ventricle, form a series of slices that (ideally) cover the entire heart. The number of SAX slices ranged from 1 to 23. Typically, the region of interest (ROI) is only a small part of the entire image.

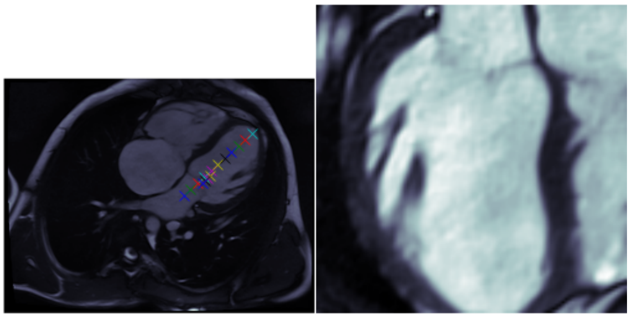

Figure 1. On the left: an example of an MRI-image in the dataset. The crosses show the intersections between the images taken over multiple axes. On the right, a cut out version of this MRI-image after focusing on the ROI using our preprocessing method.

Given a patient's data, we were asked to output a cumulative distribution function over the volume, ranging from 0 to 599 ml, for both systole and diastole. The models were scored by a Continuous Ranked Probability Score (CRPS) error metric, which computes the average squared distance between the predicted CDF and a Heaviside step function representing the real volume.

[1] http://www.datasciencebowl.com/

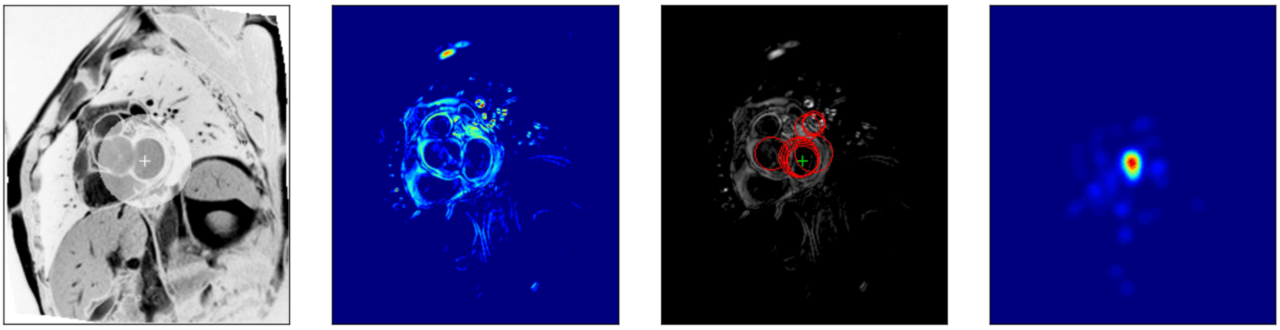

Figure 2. Our preprocessing method took the Fourier transform of the intensity of every pixel over time to extract the periodically changing pixels. In this Fourier image, it used the circle Hough transform to find the ventricles. It finally calculated a likelihood surface from which we could find the ROI around the ventricle of interest.

3. Methods

In our solution, we combined traditional image processing approaches like the one illustrated in Figure 2, which find the region of interest (ROI) in each slice, with convolutional neural networks(Krizhevsky et al., 2012), which perform the mapping from the extracted image patches to the predicted volumes. Given the very limited number of training samples, we combated overfitting by restricting our models to only combine the data sources in predefined ways.

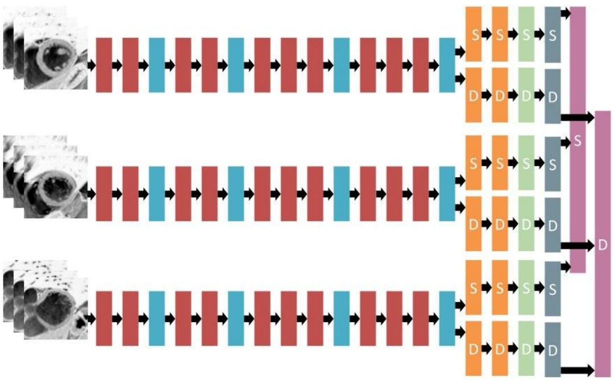

Figure 3. An illustration of one of the models in our ensemble. The red layers are convolution layers, the blue layers are maxpooling layers and the pink layer on top is a specially designed layer which does a frustum approximation between the results of the previous layers.

Finally, a key insight that led to our approach was that it would have been a bad idea to try and make one model that is capable of processing every single patient, given the large amount of inconsistencies in the data. Instead, we combined accurate models that assume clean data (SAX slices nicely ranged from one end of the heart to the other) with more robust but less accurate models to handle the outliers. One of these models is illustrated in Figure 3.

Table 1. RMS-error of various results of our approach compared to that of human annotators. (Lower is better)

	Diastolic Volume	Systolic Volume	Ejection Fraction
CNN	13.65	10.43	6.99%
Human	13	14	6%

4. Results

We finished the Kaggle competition in second place without hand-labeling either test or train data. As you may find in Table 1, our results perform equally well as human annotators, confirming the relevance of deep learning approaches for computer vision tasks, even on medical data.

Acknowledgments

The research leading to these results has received funding from the Agency for Innovation by Science and Technology in Flanders (IWT) and the Research Foundation Flanders (FWO). The Tesla K40 used for this research was donated by the NVIDIA Corporation.

References

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (pp. 1097–1105).