Journal of Information Hiding and Privacy Protection JIHPP, vol.1, no.3, pp.109-117, 2019

A Multi-Scale Network with the Encoder-Decoder Structure for CMR Segmentation

Chaoyang Xia[1], Jing Peng[1], Zongqing Ma[2] and Xiaojie Li[1, *]

Abstract: Cardiomyopathy is one of the most serious public health threats. The precise structural and functional cardiac measurement is an essential step for clinical diagnosis and follow-up treatment planning. Cardiologists are often required to draw endocardial and epicardial contours of the left ventricle (LV) manually in routine clinical diagnosis or treatment planning period. This task is time-consuming and error-prone. Therefore, it is necessary to develop a fully automated end-to-end semantic segmentation method on cardiac magnetic resonance (CMR) imaging datasets. However, due to the low image quality and the deformation caused by heartbeat, there is no effective tool for fully automated end-to-end cardiac segmentation task. In this work, we propose a multi-scale segmentation network (MSSN) for left ventricle segmentation. It can effectively learn myocardium and blood pool structure representations from 2D short-axis CMR image slices in a multi-scale way. Specifically, our method employs both parallel and serial of dilated convolution layers with different dilation rates to capture multi-scale semantic features. Moreover, we design graduated up-sampling layers with subpixel layers as the decoder to reconstruct lost spatial information and produce accurate segmentation masks. We validated our method using 164 T1 Mapping CMR images and showed that it outperforms the advanced convolutional neural network (CNN) models. In validation metrics, we archived the Dice Similarity Coefficient (DSC) metric of 78.96%.

Keywords: Cardiac magnetic resonance imaging, multi-scale, semantic segmentation, convolutional neural networks.

1 Introduction

Cardiomyopathy is one kind of serious cardiac diseases which major cause of death in the world wild. Physicians leverage cardiac magnetic resonance (CMR) images to measure the structural-functional indices, such as ventricular volume or ejection fraction, for clinical diagnosis and treatment planning. The segmentation of left ventricle (LV) on CMR images is necessary for further analysis. But cardiologists have to delineate endocardium and epicardium or other regions of clinical interest by manually. To address this time-consuming and error-prone problem, it is urgent to propose a method to accelerate and facilitate the diagnosis repeatedly and automatically. But there are a

[1] School of Computer Science, Chengdu University of Information Technology, Chengdu, 610225, China.

[2] School of Computer Science, Sichuan University, Chengdu, 610065, China.

[*] Corresponding Author: Xiaojie Li. Email: lixiaojie000000@163.com.

number of obstacles hindering automation methods into practice. For example, the shape of myocardium and ventricle is multivariate in the different cardiac cycle, the number of pixels belongs to object are extremely less than background, and the irrevocable noise appears during CMR imaging.

Many methods have been proposed past decade to solve those problems. Margeta et al. [Margeta, Geremia, Criminisi et al. (2011)] segment LV using random forests. But this method relies on image intensity and considers the segmentation task as a classification task. Moreover, the employed intensity standardization, estimation, and normalization also computationally expensive and affect the effectiveness of results. Ngo et al. [Ngo and Carneiro (2013)] use restricted Boltzmann machines (RBMs) to obtain segmentation results on a small LV dataset. But their method is semi-automated that requires user input. Liu et al. [Liu, Maere and Song (2019)] proposed an approach to find region of interest automatically, this can be used as a pre-step for segmenting organs in medical image processing. In recent years, the deep learning approach, especially CNN based method was widely applied in image segmentation. Fu et al. [Fu, Xu, Lin et al. (2016)] use deep learning model in retinal vessel segmentation. Wachinger et al. [Wachinger, Reuter and Klein (2018)] apply convolutional neural network for brain segmentation. Fu et al. [Fu, Xu, Zhang et al. (2019)] proposed a noise-resistant superpixel segmentation method for hyperspectral images segmentation. These methods design different model structures and directly learn the optimal network parameters for label prediction. The automated LV segmentation methods still need to be significantly improved.

In this work, we propose and validate a new convolutional neural network architecture for LV segmentation, which can automatically train and inference on CMR image datasets. Our proposed model is allowed to yield accurate myocardium and blood pool segmentation masks in an end-to-end way. Specifically, the architecture of our proposed model consists of two parts: encoder and decoder. The encoder uses standard 2D convolution layers repeatedly for image feature extraction, and employs depth wise separable convocation layers which used in Xception [Chollet (2017)] for effectively feature calculating and feature decoupling. Furthermore, to cope with the changeable patterns as a multi-scale object, our model stacks a serial of dilated convolution layers with different atrous rates like Chen et al. [Chen, Papandreou, Schroff et al. (2017)]. For the segmentation task, the same spatial resolution is required for segmentation mask as ground-truth. Generally, common methods restore segmentation mask spatial resolution by adding upsampling or deconvolution layers after diminished feature map. But there lacks spatial information in an abstract feature. We design the graduated upsampling layers as the decoder to supplement the vanished spatial information pass through successive convolution layers. Besides, we adopt subpixel layer [Shi, Caballero, Husz´ar et al. (2016)] to do efficient upsampling and get more accurate segmentation masks. We validated our method using 164 CMR images and showed that it outperforms the advanced convolutional neural network (CNN) models.

2 Methodology

This section presents our proposed encoder-decoder model for multi-scale LV segmentation. The base architecture of the proposed model is depicted in Fig. 1. This model takes short-axis CMR images as input. The main idea of our model design is inspired by U-Net [Ronneberger, Fischer and Brox (2015)]. There are two branches with different duties in model, encoder, and decoder. The encoder mainly responds to feature extraction and semantic encoding, and the decoder mainly responds to spatial information recovery. Skip connections links between them. To get more accurate segmentation and robustness performance, we introduce ASPP block [Chen, Papandreou, Schroff et al. (2017)] and subpixel layer into the network.

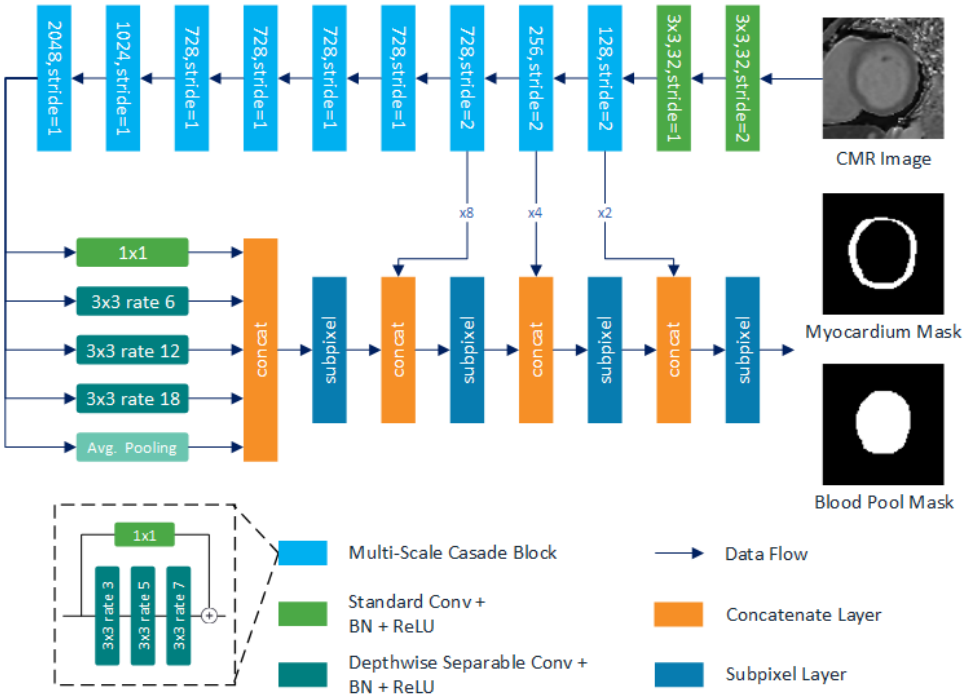

Figure 1: The architecture of our proposed MSSN network (best view in color). It consists of two parts, encoder and decoder. A blue block represents standard 1×1 convolution to reduce channel. The parallel multi-scale block uses dilated convolution layers with rates=[6, 12, 18]. The serial multi-scale block uses dilated convolution layers with rates=[3, 5, 7]. ⊕ represents residual connection

2.1 Encoder and decoder

Inspired by the successful of convolution [Krizhevsky, Sutskever and Hinton (2012); Simonyan and Zisserman (2014); He, Zhang, Ren et al. (2016)] in image recognition tasks. We use continuous convolution layers with ReLU nonlinear activation function as the encoder for image feature extraction. The decreased feature map contains more representational information, it is beneficial for an object to identification. With the

reduction of the feature map size, there exists more semantic information but less spatial information. While segmentation task needs to generate results are the same with the original image at the resolution level. Inspired by the Upsampling or Deconvolution (also called Transposed convolution layer) layers in FCN [Long, Shelhamer and Darrell (2015)], U-Net, we design a bilinear upsampling layer for space dimension recovery. But the giant upsampling gap between feature map and the scaling object will cause distortion. To get more fine segmentation mask, we use skip connection on strides=×8, ×4, ×2 level to combine fine appearance information at shallow layers and coarse semantic information at deep layers. The end segmentation result is a dense feature map predicting the probability of each pixel in the input image.

2.2 Multi-scale feature extraction

The shape of endocardial and epicardial contours is variability across different cardiac cycle phases. In order to fit object scales, some methods would crop [Zeiler and Fergus (2014)] or warp [Girshick, Donahue, Darrell et al. (2014)] apply multi-scale images sequentially from coarse-to-fine as image pyramid [Chen, Papandreou, Kokkinos et al. (2018); Chen, Yang, Wang et al. (2016)], which however leads to loss of the exact position of the segmented pixels. Inspired by Spatial Pyramid Pooling (SPP) [He, Zhang, Ren et al. (2015)] and Atrous Spatial Pyramid Pooling (ASPP) [Chen, Papandreou, Schroff et al. (2017)], we use dilated convolution layers with different rates in both serial and parallel compound manner. In this fashion, the model can capture multiscale information without multi-scale inputs. In this experiment, we revisit parallel dilated convolution layers with rates=[6, 12, 18] and three serial dilated convolution layers with rates=[3, 5, 7]. The serial dilated convolution layers with different rates and strides formed the multi-scale block.

2.3 Subpixle layer

Segmentation tasks require that the size of the segmented mask be the same as that of the original input image. Although the reduced high-level feature maps obtained by continuous downsampling operations are beneficial to classification tasks, the reduced size is not conducive to segmentation tasks. Upsampling with too large gaps will result in inadequate segmentation mask. With the success of subpixel convolutional layer in a single image and video super-resolution task, we employ the sub-pixel convolution layer in our decoder to upscale the feature maps into the original output. The critical periodic shuffling (PS) operation can be described in the following mathematical formula:

$$\mathcal{PS}(T)_{x,y,c} = T_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, C \cdot r \cdot \mathrm{mod}(y,r) + C \cdot \mathrm{mod}(x,r) + c} \tag{1}$$

where the x , y , c represent the dimensions of the feature map, and the r is scale factor.

3 Experiments and evaluation

In this section, we present the evaluation of our method on 164 short-axis T1 Mapping CMR images datasets for automated LV segmentation task. Compare our method with other advanced segmentation models on metrics of Dice Similarity Coefficient (DSC),

Jaccard Coefficient (also known as IoU) and Confidence (also called as Precision).

3.1 CMR datasets

The CMR images dataset used in our study is obtained from one hospital. All 164 images are 2D short-axis native T1 Mapping CMR images. The spacing sizes of images are ranges from 1.172×1.172×1.0 mm^3 to 1.406×1.406×1.0 mm^3. The original dimension size is 256×218×1 pixels. The ground-truth of myocardium segmentation masks for all image samples are provided by an experienced cardiologist contour manually. The CMR images in the dataset are randomly divided into training, validation and testing sets as 5-fold cross-validation.

3.2 Data pre-process

We observe that the CMR images are anisotropy in three dimensions. To have a better performance of the model, we process image data before feeding images into the model. The first step we apply on images is resampling, we use Simple-ITK software package sampling the spacing size to 1×1×1 mm^3 for isotropy. There is a wide range of pixel intensities in CMR imaging, which will directly influence the learning performance of deep learning model. Then, we normalize the pixel intensity distribution of each imaging range from -1.0 to 1.0. In MR imaging scan, there is a broad scan range on chest radiograph in MR images. For the purposes of removing unnecessary background pixels and zooming segment object, we crop each image to get our region of interest (ROI). Cropping images to reduce input spatial dimensions can also accelerate computations.

3.3 Implementation details

To have an unbiased evaluation, we use 5-fold cross validation in our experiments stage. According to the dataset, we implement LV myocardium and blood poo segmentation separately in this study. For algorithm Implementation, we use Keras neural networks library which uses TensorFlow as backend. The object function we adopted is dice loss, which optimizes the Dice Similarity Coefficient (DSC) directly. But we set dice loss as negative DSC for Keras purpose because neural networks library always minimizes the value of the loss. Moreover, the optimizer we applied is RMSprop [Tieleman and Hinton (2012)] with learning rate 0.003 and leave the other parameters of this optimizer at their default values. We use a batch size of 32 continued 40 epochs. To enhance the generalization ability of network and enlarge the size of datasets [Pan, Qin, Chen et al. (2019)], we adopt real-time data augmentation in train stage, which rotate and shift images randomly in each iteration. The model training and testing is performed on a single NVIDIA Tesla M40 GPU.

3.4 Results

To verify the performance on LV segmentation of our proposed method, we apply our model compare with other advanced segmentation methods with the same datasets and experimental settings.

The qualitative LV myocardium and LV blood pool segmentation results and the

quantitative comparison with evaluation metrics of myocardium segmentation. To visually show the segmentation effect, we overlap the original image, segmentation mask and ground-truth with different colors as Fig. 2 and Fig. 3 and make a numerical description using evaluation index in Tab. 1.

The qualitative LV myocardium and LV blood pool segmentation results are shown in Fig. 2 and Fig. 3, respectively. In the figure, each row corresponds to one image of the patient. The original T1 Mapping CMR image is displayed in the first column. To have more expressive illustration about the accuracy of automated segmentation mask, we overlap segmentation mask generated by automated algorithm on ground-truth delineated manually by cardiologist expert. The color blue represents the area of correct segmentation; the red represents the area of model unidentified which is the object, and the green represents the area of model misidentified which does not object. We compared our method with U-Net and DeepLab V3+ [Chen, Zhu, Papandreou et al. (2018)]. It is easy to observe that, the results yield by our model have larger blue (correct) areas and less red and green (wrong) areas.

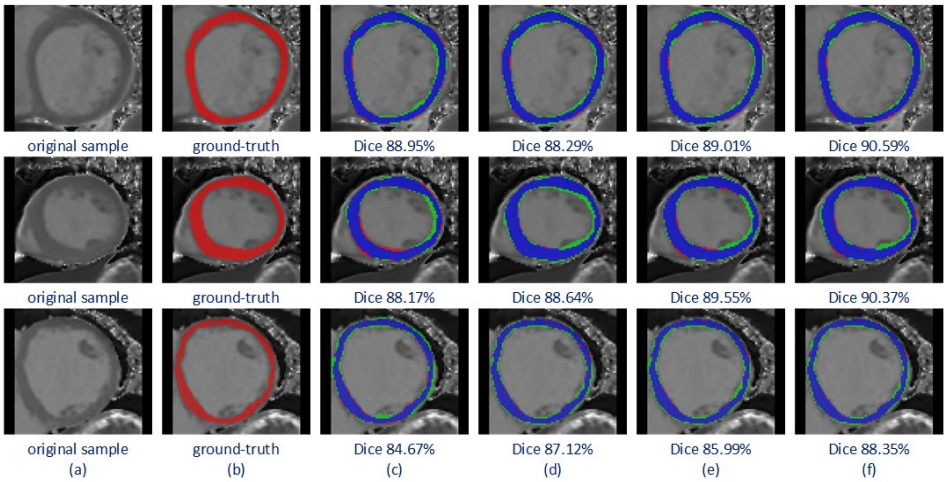

Figure 2: The qualitative myocardium segmentation results comparison (best view in color). Each row represents an example sample. The first column is the original T1 Mapping CMR image. The second column is ground-truth of LV myocardium delineated by a cardiologist. The third to sixth columns are overlapped segmentation masks with ground-truth which predicted by (c) U-Net, (d) DeepLabV3+, (e) MSSN without subpixel and (f) MSSN with subpixel, respectively. Color representation of column (c) to (f): blue: correct pixels, red: unidentified pixels, green: misidentified pixels

The quantitative comparison on evaluation metrics of myocardium segmentation is shown in Tab. 1. Evaluation metrics we including are Dice Coefficient (DSC), Jaccard Coefficient (JACC) and Confidence (CONF, also called precision) [Taha and Hanbury (2015)]. Our method archived a mean DSC value of 78.96%, Jaccard index value of 65.84% and a confidence value of 69.14%. It compared to U-Net and DeepLabV3+ of 75.11%, 61.14%, 64.09% and 77.23%, 63.59%, 66.32%, respectively.

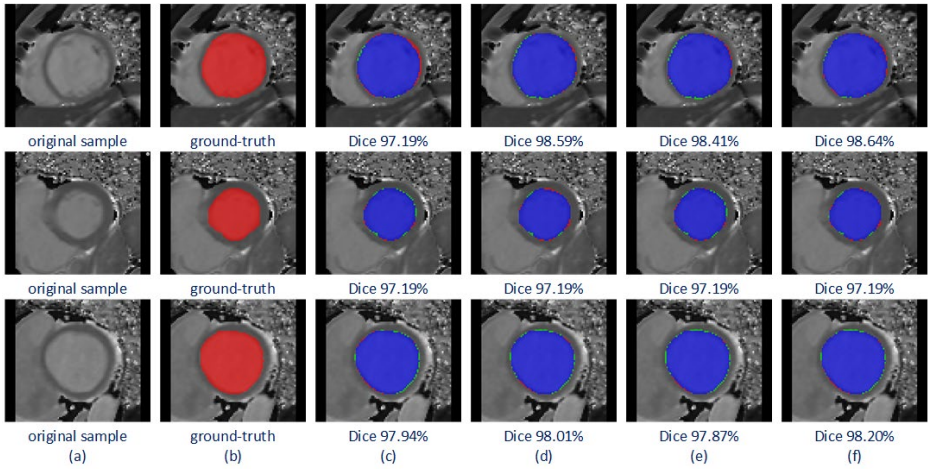

Figure 3: The qualitative blood pool segmentation results comparison (best view in color). Each row represents an example sample. The first column (a) is the original T1 Mapping CMR image. The second column (b) is ground-truth of LV blood pool delineated by a cardiologist. The third to sixth columns are overlapped segmentation masks with ground-truth which predicted by (c) U-Net, (d) DeepLabV3+, (e) MSSN without subpixel and (f) MSSN with subpixel, respectively. Color representation of column (c) to (f): blue: correct pixels, red: unidentified pixels, green: misidentified pixels

Table 1: Quantitative comparison of myocardium segmentation

Methods	DSC	JACC	CONF
U-Net	75.11%	61.14%	64.09%
DeepLabV3+	77.23%	63.59%	66.32%
MSSN	77.83%	64.49%	67.51%
MSSN + subpixel	78.96%	65.84%	69.14%

4 Conclusions

In this work, we proposed a novel end-to-end CNN model for fully automatic LV segmentation on T1 mapping CMR image datasets. The model employed serial and parallel dilated convolution layers with different dilate rates for more accuracy and robustness multi-scale myocardium extraction. And graduated subpixel upsampling layers with concatenate layers in decoder combine spatial features to from shallow layers and semantic features from deep layers. The combination of spatial information and semantic information can remedy the loss of spatial information caused by upsampling to produce more accurate segmentation masks. And the practicability and performance of our segmentation method are successfully demonstrated though evaluating on CMR image datasets and compared with other advanced segmentation methods.

Acknowledgement: This work was supported by the Project of Sichuan Outstanding Young Scientific and Technological Talents (19JCQN0003), the major Project of Education Department in Sichuan (17ZA0063 and 2017JQ0030), and in part by the Natural Science Foundation for Young Scientists of CUIT (J201704) and the Sichuan Science and Technology Program (2019JDRC0077).

References

Chen, L. C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A. L. (2018): Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848.

Chen, L. C.; Papandreou, G.; Schroff, F.; Adam, H. (2017): Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.

Chen, L. C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A. L. (2016): Attention to scale: scale-aware semantic image segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3640-3649.

Chen, L. C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. (2018): Encoderdecoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision*, pp. 801-818.

Chollet, F. (2017): Xception: deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251-1258.

Fu, H.; Xu, Y.; Lin, S.; Wong, D. W. K.; Liu, J. (2016): Deepvessel: retinal vessel segmentation via deep learning and conditional random field. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 132-139.

Fu, P.; Xu, Q.; Zhang, J.; Geng, L. (2019): A noise-resistant superpixel segmentation algorithm for hyperspectral images. *Computers, Materials & Continua*, vol. 58, pp. 509-515.

Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. (2014): Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587.

He, K.; Zhang, X.; Ren, S.; Sun, J. (2015): 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.

He, K.; Zhang, X.; Ren, S.; Sun, J. (2016): Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems.*

Liu, Z.; Maere, C.; Song, Y. (2019): Novel approach for automatic region of interest and seed point detection in ct images based on temporal and spatial data. *Computers, Materials &Continua*, vol. 58, pp. 669-686.

Long, J.; Shelhamer, E.; Darrell, T. (2015): Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.

Margeta, J.; Geremia, E.; Criminisi, A.; Ayache, N. (2011): Layered spatiotemporal forests for left ventricle segmentation from 4D cardiac MRI data. *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 109-119.

Ngo, T. A.; Carneiro, G. (2013): Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks. *IEEE International Conference on Image Processing*, pp. 695-699.

Pan, L.; Qin, J.; Chen, H.; Xiang, X.; Li, C. et al. (2019): Image augmentation-based food recognition with convolutional neural networks. *Computers, Materials & Continua*, vol. 59, no. 1, pp. 297-313.

Ronneberger, O.; Fischer, P.; Brox, T. (2015): U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241.

Shi, W.; Caballero, J.; Husz'ar, F.; Totz, J.; Aitken, A. P. et al. (2016): Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874-1883.

Simonyan, K.; Zisserman, A. (2014): Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Taha, A. A.; Hanbury, A. (2015): Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, vol. 15, no. 1, pp. 29.

Tieleman, T.; Hinton, G. (2012): Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, vol. 4, no. 2, pp. 26-31.

Wachinger, C.; Reuter, M.; Klein, T. (2018): Deepnat: deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, vol. 170, pp. 434-445.

Zeiler, M. D.; Fergus, R. (2014): Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, pp. 818-833.