Automated Multi-sequence Cardiac MRI Segmentation Using Supervised Domain Adaptation

Sulaiman Vesal[1]([✉]), Nishant Ravikumar[1,2], and Andreas Maier[1]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[2] CISTIB, Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing, LICAMM Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, LS2 9JT, United Kingdom

`sulaiman.vesal@fau.de`

Abstract. Left ventricle segmentation and morphological assessment are essential for improving diagnosis and our understanding of cardiomyopathy, which in turn is imperative for reducing risk of myocardial infarctions in patients. Convolutional neural network (CNN) based methods for cardiac magnetic resonance (CMR) image segmentation rely on supervision with pixel-level annotations, and may not generalize well to images from a different domain. These methods are typically sensitive to variations in imaging protocols and data acquisition. Since annotating multi-sequence CMR images is tedious and subject to inter- and intra-observer variations, developing methods that can automatically adapt from one domain to the target domain is of great interest. In this paper, we propose an approach for domain adaptation in multi-sequence CMR segmentation task using transfer learning that combines multi-source image information. We first train an encoder-decoder CNN on T2-weighted and balanced-Steady State Free Precession (bSSFP) MR images with pixel-level annotation and fine-tune the same network with a limited number of Late Gadolinium Enhanced-MR (LGE-MR) subjects, to adapt the domain features. The domain-adapted network was trained with just four LGE-MR training samples and obtained an average Dice score of ∼85.0% on the test set comprises of 40 LGE-MR subjects. The proposed method significantly outperformed a network without adaptation trained from scratch on the same set of LGE-MR training data.

Keywords: Multi-sequence MRI · Deep Learning · Domain Adaptation · Myocardial Infraction · MRI segmentation

1 Introduction

Myocardial infarction (MI) is the leading cause of mortality and morbidity worldwide[1][2]. Accurate analysis and modeling of the ventricles and myocardium from medical images are essential steps for diagnosis and treatment of patients

with MI [3]. MR imaging is used in the clinical workflow to provide anatomical and functional information of the heart. Different types of CMR sequences are acquired to provide complimentary information to each other, for example, T2-weighted images highlight the acute injury and ischemic regions, and the bSSFP cine sequence captures cardiac motion and presents clear boundaries. Moreover, LGE CMR can enhance the infarcted myocardium, appearing with distinctive brightness compared with healthy tissue [4]. It is widely used to study the presence, location, and extent of MI in clinical studies. Thus, segmenting ventricles and myocardium from LGE CMR images is important to predict risk of infarcts, identify the extent of infarcted tissue and for patient prognosis [5]. However, manual delineation is generally time-consuming, tedious and subject to inter- and intra-observer variations [6]. In the medical image domain, heterogeneous domain shift is a severe problem, given the diversity in imaging modalities. For example, as shown in Fig. 1, cardiac regions visually appear significantly different in images acquired using different MR sequences. Generally, deep learning models trained on one set of MR sequence images perform poorly when tested on another type of MR sequence. One approach to maintain model performance in such a setting is to employ domain adaptation e.g. image to image translation or transfer learning. Domain adaptation attempts to reduce the shift between the distribution of data within the source and target domain.

Related work. Existing methods have approached multi-modal CMR segmentation using techniques such as cross-constrained shape [7], generative adversarial networks or 3D CNN. In [8][9], the authors first trained a CNN on the source domain and then transformed the target domain images into the appearance of source images, such that they could be analyzed using the network pre-trained on the source domain. However, these methods are based on generative adversarial networks and required substantial training data to achieve stable performance. On the other hand, there are limited works focusing on automatic LGE-CMR segmentation, which is a crucial prerequisite in a number of clinical applications of cardiology. Recently, few studies have attempted CMR multi-sequence segmentation. This type of method uses complementary information from multiple sequences to segment heart structures. [6] proposed an unsupervised method using a multivariate mixture model (MvMM) for multi-sequence segmentation. MvMM adopted to model the joint intensity distribution of the the multi-sequence images. The performance of this method depends on the quality of registration.

Contributions. In this study, we develop a deep learning-based method to segment the ventricles and myocardium in LGE CMR, combined with two other sequences (T2 and bSSFP) from the same patients. T2 and bSSFP sequences are used to assist the LGE CMR segmentation. Our method introduces a feature adaptation mechanism using transfer learning which explicitly adapts the features from T2 and bSSFP sequences to LGE images with few target training data. We first train an encoder-decoder CNN on T2 and bSSFP images with pixel-level annotation and re-train the same network with a limited number of

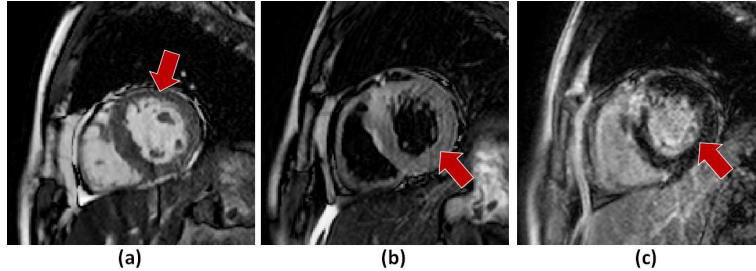LGE images, to adapt the learned features and imbue domain invariance between source and target domains.

Fig. 1. Illustration of different CMR sequences: a) bSSFP slice, b) T2-weighted slice and c) LGE slice. The red arrows point to left ventricle on different sequences.

2 Method

2.1 Domain Adaptation

Deep learning methods are typically sensitive to domain shift and perform poorly on a new set of data with a different marginal probability distribution. However, annotating data for every new domain is a very expensive task, particularly in the medical area that requires clinical expertise. To segment LGE CMR images with very few annotated subjects, we attempted to get complementary information from other sequences with pixel-level annotations, and transfer the domain knowledge and initialize a second network with pre-trained weights. This is called as supervised domain adaptation [10].

Let's consider D_{tb} as the image domain for T2+bSSFP sequences and D_l for LGE sequences respectively. D_{tb} can be expressed with feature space S and associated probability distribution of $P(X)$ where $X = \{x_1, x_2, ..., x_n\} \in S$ [11][12]. In a supervised learning task, domain $D_{tb} = \{S, P(X)\}$ consists of a model with associated objective function $\mathcal{F}_{tb}$, learning task of T_{tb} and a label space of Y. The objective function $\mathcal{F}_{tb}$ for task segmentation T_{tb} can be optimised using a pair of samples $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$. After the training process, the learned model $\hat{\mathcal{F}}_{tb}$ can be used to predict on new samples of T2 and bSSFP images from D_{tb} domain. Now, if we consider D_l with LGE segmentation task T_l, we can transfer the learned weights from domain D_{tb} to improve objective function of F_l for segmenting LGE images in domain D_l where $D_{tb} \neq D_l$ and $T_{tb} \neq T_l$. In this way, domain D_l uses information from domain D_{tb} to segment LGE images. The final model trained on T2+bSSFP domain and adapted to target domain (LGE) can be denoted as $\mathcal{F}_{lge}$.

To construct the model $\mathcal{F}_{lge}$, we transferred the learned weights from $\mathcal{F}_{tb}$, then we retrain all the layers and fine-tuned the model on the limited training

data from domain D_l. This is demonstrated in Fig. 2. All hyperparameters associated with the optimizer, the loss function, and the data augmentation scheme employed were kept the same for both models.
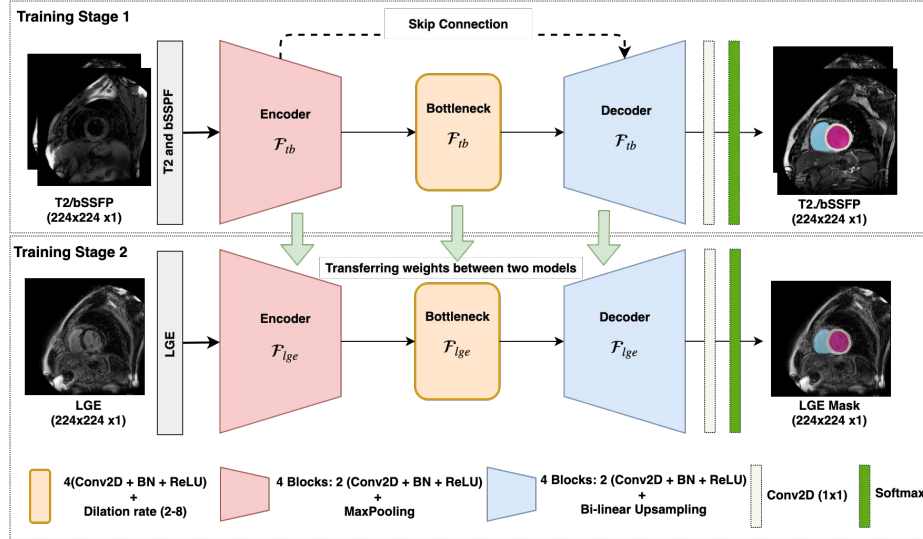

Fig. 2. Overview of our network architecture for feature transfer learning between different CMR sequences. The encoder-decoder first trained with T2+bSSFP images and in the second stage the network with learned weights retrained with LGE iamges.

2.2   Network Architecture

Our network architecture is a fully convolutional network inspired from [13] which comprises four encoder and decoder blocks, separated by a bottleneck block (refer to Fig. 2). The architecture includes skip connections between all encoder and decoder blocks at the same spatial resolution. Each encoder/decoder block consists of two 2D convolution layers, where, each convolution layer is followed by a batch-normalization and a Rectifier Linear Unit (ReLU) layer. In each encoder-convolution block, the input of the first convolution layer is concatenated with the output of the second convolution layer and zero-padded accordingly. The subsequent 2D max-pooling layer reduces the dimensions of the image by half. The use of residual connections [14] between convolution layers of each block in the encoder, help improve the flow of gradients in the backward pass of the network. The network utilizes a 1×1 convolution to aggregate the feature maps from the final decoder block. This operation improves discriminative power as feature maps with lower activations are more likely to be suppressed through the assignment of lower weights. Finally, a softmax activation function

was used in the last layer of the first network to classify the background from the foreground classes. Compared to U-Net, we replace the bottleneck convolution layers of the network with dilated convolutions [15] of size 3×3, to enlarge the receptive field and enable the network to capture both local and global contextual information. The dilation rate of the four convolution layers is increased successively from $1 - 8$, and subsequently, their feature maps are summed together, enabling the network to capture the entire image's field of view.

Multi-class Dice loss: To train the proposed network, a modified version of the soft-Dice loss is used which is less sensitive to class imbalance. This is motivated by the successful recent works [13][16] for medical image segmentation. The Dice score is computed for each class individually, and then averaged over the number of classes. In order to segment an $N \times N$ input image (for example, a T1-weighted image with LV, RV, Myo and background as labels), the output of softmax layer is four probabilities for classes $k = 0, 1, 2, 3$ where, $\sum_c y_{n,k} = 1$ for each pixel. Given the one-hot encoded ground truth label $\hat{y}_{n,k}$ for that corresponding pixel, the multi-class soft Dice loss is defined as follows:

$$\zeta_{dc}(y, \hat{y}) = 1 - \frac{1}{K}\left(\sum_k \frac{\sum_n y_{nk}\hat{y}_{nk}}{\sum_n y_{nk} + \sum_n \hat{y}_{nk}}\right) \tag{1}$$

2.3   Data Acquisitions

We validated our proposed method on the STACOM MS-CMRSeg[3] 2019 challenge dataset with short-axis cardiac MR images of 45 patients diagnosed with cardiomyopathy. The dataset was collected in Shanghai Renji hospital with institutional ethics approval [6]. Each patient had been scanned using three CMR sequences: LGE, T2, and bSSFP. Ground truth masks of cardiac structures were provided for 35 training samples (T2 and bSSFP only) and 5 validation samples, including the Left ventricle cavity (LV), the right ventricle cavity (RV), and the myocardium of the left ventricle (Myo). The LGE CMR was a T1-weighted, inversion-recovery, gradient-echo sequence, consisting of 10 to 18 slices covering the main body of the ventricles. The acquisition matrix was 512×512, yielding an in-plane resolution of 0.75×0.75 mm and slice thickness of 5 mm. The bSSFP CMR images consist of 8 to 12 contiguous slices, covering the full ventricles from the apex to the basal plane of the mitral valve, with some cases having several slices beyond the ventricles. These sequences have a slice thickness of 8-13 mm and an image resolution of 1.25×1.25 mm. The T2-weighted CMR images consist of a small number of slices. Few cases have comprise just three slices, and the others have five (13 subjects), six (8 subjects) or seven (one subject) slices. The slice thickness is 12-20 mm and an in-plane resolution of 1.35×1.35 mm. Since T1 and bSSFP images have very few slices, we combined both sequences together for the training of our backbone network.

Preprocessing: There is a large degree of variance in contrast and brightness across the MS-CMRSeg 2019 challenge images. The variability results from

---

[3] http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mscmrseg19/

different system settings, and data acquisition which makes it harder for neural networks to process the images. Due to low contrast, we enhanced the image contrast slice-by-slice, using contrast limited adaptive histogram equalization (CLAHE). We normalized each MR volume individually to have zero mean and unit variance and cropped all images to 224×224 to remove the black areas(background regions). Fig. 1 shows MR slices of three different sequences of a patient after prepossessing. Furthermore, we use common training data augmentation strategies including random rotation, random scaling, random elastic deformations, random flips, and small shifts in intensity to increase training data. We employed the augmentation only on x and y axes and kept the volume depth the same. In this way, we do not degrade image quality.

2.4   Network Training

The organizers of the MS-CMRSeg challenge already split the data into training, validation and testing sets. In the first stage, we trained our model on 35 T2 and bSSFP sequences with pixel-level annotations and validated on 5 subjects using the adaptive moment estimation (ADAM) optimizer. For the second stage, we re-trained the model with five LGE subjects using 5-fold cross validation. The learning rate was fixed at 0.0001, and the exponential decay rates of the 1st and 2nd-moment estimates were set to 0.9 and 0.999, respectively. During training, segmentation accuracy was evaluated on the validation set after each epoch of the network. Networks were trained until the validation accuracy stopped increasing, and the best performing model was selected for evaluation on the test set. The batch size for training T2+bSSFP images was set to 16 and during fine-tuning for LGE-MRI, to 4. We employed connected component (CC) analysis as a post-processing step to remove the small miss-classified regions in the output of the softmax layer. For inference, we use the weights which achieved the best Dice score on the validation set of LGE-CMR data set. The network was developed in Keras and TensorFlow, an open-source deep learning library for Python, and was trained on an NVIDIA Titan X-Pascal GPU with 3840 CUDA cores, and 12GB RAM.

2.5   Evaluation Criteria

To evaluate the accuracy of segmentation results, we used three different metrics to evaluate segmentation accuracy, namely, the Dice coefficient (Dice), Hausdorff distance (HD), and Average surface distance (ASD). The Dice metric measures the degree of overlap between the predicted and ground truth segmentation. It is the most widely used metric for evaluating segmentation quality in medical imaging. HD is defined as the maximum of the minimum voxel-wise distances between the ground truth and predicted object boundaries. ASD is the average of the minimum voxel-wise distances between the ground truth and predicted object boundaries. HD and ASD are evaluated using the shortest Euclidean distance of an arbitrary voxel v to a point P, defined as $\bar{d}(v, P) = \min_{p \in P} ||v - p||$.

3 Results and Discussion

The proposed model is evaluated on the task of LGE-MRI segmentation. We compare our domain adapted network with three different training strategies including training without domain adaptation and predicting using the model trained on T2 and bSSFP only. Table 1. summarizes the comparison results, where we can see that our proposed method significantly improved the segmentation performance relative to networks tested without the adaptation strategy, in terms of Dice, HD, and ASD metrics. The model without domain adaptation and trained from scratch only on 5 LGE MR samples achieved an average Dice of 66.9% on the validation set. Remarkably, for our proposed network with data augmentation, the average Dice improved to 80.9% and HD and ASD was reduced to 13.6 and 1.0mm respectively. We achieved over 87.1% Dice score for the LV structure and over 80.2% Dice score for the RV. To illustrate the domain shift problem, we directly feed LGE-MR images to the first encoder-decoder after supervised training on T2+bSSFP domain. The result in Table 1 indicates that the source model completely failed on LGE-MR images with an average Dice score of 31.3%, HD score of 37.37mm and ASD value of 11.06mm. Notably, compared with testing using model trained on T2+bSSFP, our method achieved superior performance especially for the LV and Myo structures, which are difficult to segment due to the presence of scars and blood pool within the cavity. Fig. 3 illustrates segmentation results produced with different methods. Our method demonstrated robust segmentation performance on 40 LGE CMR sequences in the test dataset, summarized in Table 2. We have achieved an average Dice score of 84.5% and HD, ASD score of 13.6 and 2.2mm respectively. The proposed model is able to reach a Dice score of 0.788 in terms of myocardium segmentation. These results help to highlight the generalization capacity of our approach to segmenting cardiac structures in LGE-MRI. Adapting the features between two domains results to a better model weight initialization and consequently improved the discrimination power of the second model.

Table 1. Performance comparison between our proposed method and other segmentation methods for LGE-MR image segmentation on the validation data set.

Methods	Dice ↑				HD [mm] ↓				ASD [mm] ↓			
	Myo	LV	RV	Avg	Myo	LV	RV	Avg	Myo	LV	RV	Avg
W/o adaptation	0.527	0.775	0.705	0.669	57.23	129.67	124.13	103.68	2.96	1.37	1.35	1.89
T2+bSSFP	0.169	0.386	0.383	0.313	42.62	30.37	39.13	37.37	14.6	7.04	11.53	11.06
W-adaptation	0.671	0.862	0.766	0.766	17.6	13.25	22.78	17.88	2.77	**0.75**	3.81	2.44
W-adaptation+Aug	**0.749**	**0.871**	**0.802**	**0.807**	**11.35**	15.66	14.05	**13.69**	**1.06**	0.81	**1.18**	**1.02**

4 Conclusion

In this study, we developed a robust deep learning approach for multi-sequence CMR image segmentation based on feature/domain-adaptation. Our network
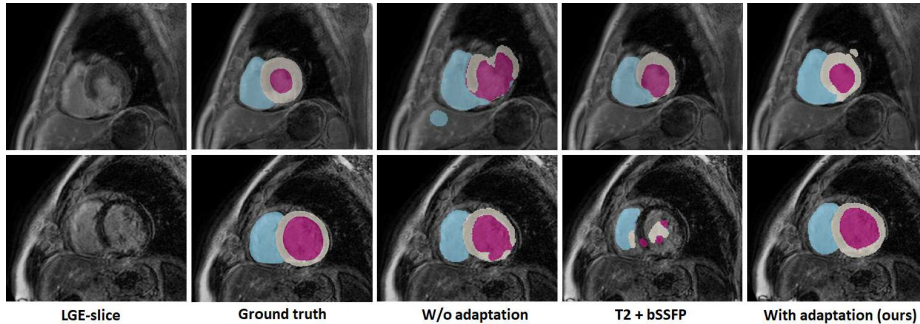
LGE-slice   **Ground truth**   **W/o adaptation**   **T2 + bSSFP**   **With adaptation (ours)**

Fig. 3. Visual comparison of segmentation results produced by different methods. From left to right are the raw LGE-MR images, ground truth, W/o Adaptation output trained from scratch on LGE-MR images only, output from model trained on T2+bSSFP images, and proposed network. The cardiac structures of LV, RV, and Myo are indicated in purple, blue and withe color respectively. Each row corresponds to one subject.

Table 2. The Dice, Jaccard, HD and ASD score of domain-adapted method on the LGE CMR test dataset.

Structure	Test-set results ($m \pm sd$)			
	Dice ↑	**Jaccard** ↑	**ASD** [mm] ↓	**HD** [mm]↓
Myo	0.788 ± 0.073	0.656 ± 0.096	2.036 ± 0.616	12.53 ± 3.37
LV	0.912 ± 0.033	0.840 ± 0.056	1.806 ± 0.615	11.29 ± 4.55
RV	0.832 ± 0.084	0.721 ± 0.117	2.804 ± 1.376	17.11 ± 6.14
Average	$\mathbf{0.844 \pm 0.063}$	$\mathbf{0.740 \pm 0.090}$	$\mathbf{2.215 \pm 0.869}$	$\mathbf{13.64 \pm 4.68}$

was first trained on T2-weighted and bSSFP sequences, and subsequently, the learned model weights were used to initialize the second network, and fine-tuned to segment LGE-MR images with a limited number of samples. The transfer leaning mechanism drastically reduces the domain shift during the training process. We validated our method on multi-sequence CMR images by comparing it with networks trained without domain adaptation. We employ our network on 2D slices as 3D models did not perform well on the MS-CMRSeg challenge 2019 dataset, given the limited number of axial slices. Experimental results highlight the advantage afforded by our approach, with regards to segmentation accuracy in LGE-MRI. Future work will aim to extend the framework using techniques for joint unsupervised image and feature adaptation using generative adversarial networks.

References

1. Kim, H.W., Farzaneh-Far, A., Kim, R.J.: Cardiovascular magnetic resonance in patients with myocardial infarction: Current and emerging applications. Journal of the American College of Cardiology **55**(1) (2009) 1 – 16
2. Group, E.S.D., Timmis, A., Wilkins, E., Wright, L., Townsend, N., Gale, C., Grobbee, R., Vos, R., Blum, M., Maniadakis, N., Vardas, P., Flather, M., Bax, J., Pinto, F.: European Society of Cardiology: Cardiovascular Disease Statistics 2017. European Heart Journal **39**(7) (11 2017) 508–579
3. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019) 1–1
4. Hammer-Hansen, S., Bandettini, W.P., Hsu, L.Y., Leung, S.W., Shanbhag, S., Mancini, C., Greve, A.M., Kber, L., Thune, J.J., Kellman, P., Arai, A.E.: Mechanisms for overestimating acute myocardial infarct size with gadolinium-enhanced cardiovascular magnetic resonance imaging in humans: a quantitative and kinetic study. European Heart Journal - Cardiovascular Imaging **17**(1) (05 2015) 76–84
5. Kurzendorfer, T., Forman, C., Schmidt, M., Tillmanns, C., Maier, A., Brost, A.: Fully automatic segmentation of left ventricular anatomy in 3-DLGE-MRI. Computerized Medical Imaging and Graphics (59) (2017) 13–27
6. Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence mri. In: MICCAI. (2016) 581–588
7. Liu, J., Xie, H., Zhang, S., Gu, L.: Multi-sequence myocardium segmentation with cross-constrained shape and neural network-based initialization. Computerized Medical Imaging and Graphics **71** (2019) 49 – 57
8. Chen, C., Dou, Q., Chen, H., Heng, P.A.: Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation. In: Machine Learning in Medical Imaging. (2018) 143–151
9. Russo, P., Carlucci, F.M., Tommasi, T., Caputo, B.: From source to target and back: Symmetric bi-directional adaptive gan. In: CVPR. (June 2018)
10. Motiian, S., Piccirilli, M., Adjeroh, D.A., Doretto, G.: Unified deep supervised domain adaptation and generalization. In: 2017 IEEE International Conference on Computer Vision (ICCV). (Oct 2017) 5716–5726
11. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10) (Oct 2010) 1345–1359
12. Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R.G., de Leeuw, F.E., Tempany, C.M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B., Wells, W.M.: Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In: MICCAI 2017. (2017) 516–524
13. Vesal, S., Ravikumar, N., Maier, A.: Dilated convolutions in neural networks for left atrial segmentation in 3d gadolinium enhanced-mri. In: STACOM. (2018) 319–328
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. (June 2016) 770–778
15. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: ICLR. (2016)
16. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). (Oct 2016) 565–571