Left Ventricle Segmentation and Quantification from Cardiac Cine MR Images via Multi-task Learning

Shusil Dangi[1(✉)], Ziv Yaniv[2,3], and Cristian A. Linte[1,4]

[1] Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA
{sxd7257,calbme}@rit.edu
[2] TAJ Technologies Inc., Bloomington, MN, USA
[3] National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
[4] Biomedical Engineering, Rochester Institute of Technology, Rochester, NY, USA

Abstract. Segmentation of the left ventricle and quantification of various cardiac contractile functions is crucial for the timely diagnosis and treatment of cardiovascular diseases. Traditionally, the two tasks have been tackled independently. Here we propose a convolutional neural network based multi-task learning approach to perform both tasks simultaneously, such that, the network learns better representation of the data with improved generalization performance. Probabilistic formulation of the problem enables learning the task uncertainties during the training, which are used to automatically compute the weights for the tasks. We performed a five fold cross-validation of the myocardium segmentation obtained from the proposed multi-task network on 97 patient 4-dimensional cardiac cine-MRI datasets available through the STACOM LV segmentation challenge against the provided gold-standard myocardium segmentation, obtaining a Dice overlap of 0.849 ± 0.036 and mean surface distance of 0.274 ± 0.083 mm, while simultaneously estimating the myocardial area with mean absolute difference error of 205 ± 198 mm^2.

1 Introduction

Magnetic Resonance Imaging (MRI) is the preferred imaging modality for non-invasive assessment of cardiac performance, thanks to its lack of ionizing radiation, good soft tissue contrast, and high image quality. Cardiac contractile function parameters such as systolic and diastolic volumes, ejection fraction, and myocardium mass are good indicators of cardiac health, representing reliable diagnostic value. Segmentation of the left ventricle (LV) allows us to compute these cardiac parameters, and also to generate high quality anatomical models for surgical planning, guidance, and regional analysis of the heart. Although manual delineation of the ventricle is considered as the gold-standard, it is time consuming and highly susceptible to inter- and intra-observer variability. Hence, there is a need for fast, robust, and accurate semi- or fully-automatic segmentation algorithms.

Cardiac MR image segmentation techniques can be broadly classified into: (1) no-prior based methods, such as thresholding, edge-detection and linking, and region growing; (2) weak-prior based methods, such as active contours (snakes), level-set, and graph-theoretical models; (3) strong-prior based methods, such as active shape and appearance models, and atlas-based models; and (4) machine learning based methods, such as per pixel classification and convolutional neural network (CNN) based models. A comprehensive review of various cardiac MR image segmentation techniques can be found in [1].

Recent success of deep learning techniques [2] in high level computer vision, speech recognition, and natural language processing applications has motivated their use in medical image analysis. Although the early adoption of deep learning in medical image analysis encountered various challenges, such as the limited availability of medical imaging data and associated costly manual annotation, those challenges were circumvented by patch-based training, data augmentation, and transfer learning techniques [3,4].

Long *et al.* [5] were the first to propose a fully convolutional network (FCN) for semantic image segmentation by adapting the contemporary classification networks fine-tuned for the segmentation task, obtaining state-of-the-art results. Several modifications to the FCN architecture and various post-processing schemes have been proposed to improve the semantic segmentation results as summarized in [6]. Notably, the U-Net architecture [7] with data augmentation has been very successful in medical image segmentation.

While segmentation indirectly enables the computation of various cardiac indices, direct estimation of these quantities from low-dimensional representation of the image have also been proposed in the literature [8–10]. However, these methods are less interpretable and the correctness of the produced output is often unverifiable, potentially limiting their clinical adoption.

Here we propose a CNN based multi-task learning approach to perform both LV segmentation and cardiac indices estimation simultaneously, such that these related tasks regularize the network, hence improving the network generalization performance. Furthermore, our method increases the interpretablity of the output cardiac indices, as the clinicians can infer its correctness based on the quality of produced segmentation result.

2   Methodology

Traditionally, the segmentation of the LV and quantification of the cardiac indices have been performed independently. However, due to a close relation between the two tasks, we identified that learning a CNN model to perform both tasks simultaneously is beneficial in two ways: (1) it forces the network to learn features important for both tasks, hence, reducing the chances of overfitting to a specific task, improving generalization; (2) the segmentation results can be used as a proxy to identify the reliability of the obtained cardiac indices, and also to perform regional cardiac analysis and surgical planning.

2.1   Data Preprocessing and Augmentation

This study employed 97 de-identified cardiac MRI image datasets from patients suffering from myocardial infarction and impaired LV contraction available as a part of the STACOM Cardiac Atlas Segmentation Challenge project [11,12] database[1]. Cine-MRI images in short-axis and long-axis views are available for each case. The semi-automated myocardium segmentation provided with the dataset served as gold-standard for assessing the proposed segmentation technique. The dataset was divided into 80% training and 20% testing for five-fold cross-validation.

The physical pixel spacing in the short-axis images ranged from 0.7031 to 2.0833 mm. We used SimpleITK [13] to resample all images to the most common spacing of 1.5625 mm along both x- and y-axis. The resampled images were center cropped or zero padded to a common resolution of 192×192 pixels. We applied two transformations, obtained from the combination of random rotation and translation (by maximum of half the image size along x- and y-axis), to each training image for data augmentation.

2.2   Multi-task Learning Using Uncertainty to Weigh Losses

We estimate the task-dependent uncertainty [14] for both myocardium segmentation and myocardium area regression via probabilistic modeling. The weights for each task are determined automatically based on the task uncertainties learned during the training [15].

For a neural network with weights $\mathbf{W}$, let $\mathbf{f^W(x)}$ be the output for the corresponding input $\mathbf{x}$. We model the likelihood for segmentation task as the squashed and scaled version of the model output through a softmax function:

$$p(\mathbf{y}|\mathbf{f^W(x)}, \sigma) = \text{Softmax}\left(\frac{1}{\sigma^2}\mathbf{f^W(x)}\right) \tag{1}$$

where, σ is a positive scalar, equivalent to the *temperature*, for the defined Gibbs/Boltzmann distribution. The magnitude of σ determines the *uniformity* of the discrete distribution, and hence relates to the uncertainty of the prediction. The log-likelihood for the segmentation task can be written as:

$$\log p(\mathbf{y} = c|\mathbf{f^W(x)}, \sigma) = \frac{1}{\sigma^2}f_c^{\mathbf{W}}(\mathbf{x}) - \log \sum_{c'} \exp\left(\frac{1}{\sigma^2}f_{c'}^{\mathbf{W}}(\mathbf{x})\right) \tag{2}$$

where $f_c^{\mathbf{W}}(\mathbf{x})$ is the c'th element of the vector $\mathbf{f^W(x)}$.

Similarly, for the regression task we define our likelihood as a Lapacian distribution with its mean given by the neural network output:

$$p(\mathbf{y}|\mathbf{f^W(x)}, \sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|\mathbf{y} - \mathbf{f^W(x)}|}{\sigma}\right) \tag{3}$$

---

[1]

The log-likelihood for regression task can be written as:

$$\log p(\mathbf{y}|\mathbf{f^W}(\mathbf{x}), \sigma) \propto -\frac{1}{\sigma}|\mathbf{y} - \mathbf{f^W}(\mathbf{x})| - \log\sigma \tag{4}$$

where σ is the neural networks observation noise parameter—capturing the noise in the output.

For a network with two outputs: continuous output $\mathbf{y}_1$ modeled with a Laplacian likelihood, and a discrete output $\mathbf{y}_2$ modeled with a softmax likelihood, the joint loss $\mathcal{L}(\mathbf{W}, \sigma_1, \sigma_2)$ is given by:

$$\begin{aligned}
\mathcal{L}(\mathbf{W}, \sigma_1, \sigma_2) &= -\log p(\mathbf{y}_1, \mathbf{y}_2 = c|\mathbf{f^W}(\mathbf{x})) \\
&= -\log \left(p(\mathbf{y}_1|\mathbf{f^W}(\mathbf{x}), \sigma_1) \cdot p(\mathbf{y}_2 = c|\mathbf{f^W}(\mathbf{x}), \sigma_2) \right) \\
&\approx \frac{1}{\sigma_1}\mathcal{L}_1(\mathbf{W}) + \frac{1}{\sigma_2^2}\mathcal{L}_2(\mathbf{W}) + \log\sigma_1 + \log\sigma_2
\end{aligned} \tag{5}$$

where $\mathcal{L}_1(\mathbf{W}) = |\mathbf{y_1} - \mathbf{f^W}(\mathbf{x})|$ is the mean absolute distance (MAD) loss of $\mathbf{y}_1$ and $\mathcal{L}_2(\mathbf{W}) = -\log \text{Softmax}(\mathbf{y}_2, \mathbf{f^W}(\mathbf{x}))$ is the cross-entropy loss of $\mathbf{y}_2$. To arrive at (5), the two tasks are assumed independent and simplifying assumptions have been made for the softmax likelihood, resulting in a simple optimization objective with improved empirical results [15]. During the training, the joint likelihood loss $\mathcal{L}(\mathbf{W}, \sigma_1, \sigma_2)$ is optimized with respect to $\mathbf{W}$ as well as σ_1, σ_2.

As observed in (5), the uncertainties (σ_1, σ_2) learned during the training are weighting the losses for individual tasks, such that, the task with higher uncertainty is weighted less and vice versa. Furthermore, the uncertainties can't become too large, as they are penalized by the last two terms in (5).

2.3   Network Architecture

In this work, we adapt the U-Net architecture [7], highly successful in medical image segmentation, to perform an additional task of myocardium area estimation as shown in Fig. 1. The segmentation and regression paths are split at the final up-sampling and concatenation layer. The final feature map in the segmentation path is passed through a sigmoid layer to obtain a per-pixel image segmentation. Similarly, the regression output is obtained by down-sampling the final feature map in the regression path by $1/4^{th}$ of its size and passing it through a fully-connected layer. The logarithm of the task uncertainties ($\log\sigma_1$, $\log\sigma_2$) added as the network parameters are used to construct the loss function (5), and are learned during the training. Note that we train the network to predict the log uncertainty $s = \log(\sigma)$ due to its numerical stability and the positivity constraint imposed on the computed uncertainty via exponentiation, $\sigma = \exp(s)$.

3   Results

The network was initialized with the *Kaiming uniform* [16] initializer and trained for 50 epochs using *RMS prop* optimizer with a learning rate of 0.001 (decayed by 0.95 every epoch) in PyTorch[2]. The best performing network, in terms of the Dice overlap between the obtained and gold-standard segmentation, in the test set, was saved and used for evaluation.

The network training required 9 min per epoch on average using a 12GB Nvidia Titan Xp GPU. It takes 0.663 milliseconds on average to process a slice during testing. The log uncertainties learned for the segmentation and regression tasks during training are -3.9 and 3.45, respectively, which correspond to weighting the cross-entropy and mean absolute difference (MAD) loss by a ratio of 1556:1. Note that the scale for cross-entropy loss is 10^{-2}, whereas that for MAD loss is 10^{2}.
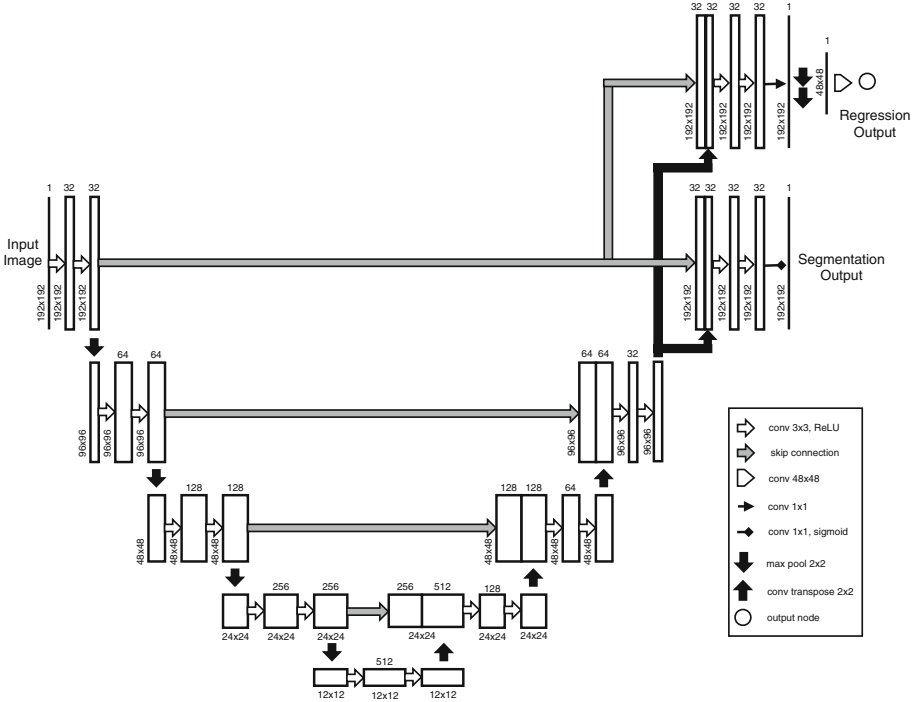

Fig. 1. Modified U-Net architecture for multi-task learning. The segmentation and regression tasks are split at the final up-sampling and concatenation layer. The final feature map in the segmentation path is passed through a sigmoid layer to obtain a per-pixel image segmentation. Similarly, the final feature map in the regression path is down-sampled (by max-pooling) to $1/4^{th}$ of its size and fed to a fully-connected layer to generate a single regression output. The logarithm of the task uncertainties ($\log\sigma_1$, $\log\sigma_2$) are set as network parameters and are encoded in the loss function (5), hence learned during the training.

---

[2] https://github.com/pytorch/pytorch.

Table 1. Evaluation of the segmentation results obtained from the baseline U-Net (UNet) architecture and the proposed multi-task network (MTN) against the provided gold-standard myocardium segmentation using—Dice Index, Jaccard Index, Mean Surface Distance, and Hausdorff Distance.

Assessment metric	End-diastole		End-systole		All phases	
	UNet	**MTN**	UNet	**MTN**	UNet	**MTN**
Dice Index	0.836 ± 0.036	**0.837 ±** 0.038	**0.850 ±** 0.033	0.849 ± 0.036	0.847 ± 0.035	**0.849 ±** 0.036
Jaccard Index	0.719 ± 0.052	**0.721 ±** 0.054	**0.740 ±** 0.048	0.739 ± 0.053	0.736 ± 0.050	**0.739 ±** 0.053
Mean surface distance (mm)	0.318 ± 0.089	**0.286 ±** 0.087	0.299 ± 0.095	**0.274 ±** 0.090	0.305 ± 0.088	**0.274 ±** 0.083
Hausdorff distance (mm)	13.582± 4.337	**13.364±** 4.108	**13.083±** 3.630	13.355± 3.861	**13.211±** 4.212	13.233± 3.810

The 2D segmentation results are stacked to form a 3D volume, and the largest connected component is selected as the final myocardium segmentation. The myocardium segmentation obtained for end-diastole, end-systole, and all cardiac phases from the proposed multi-task network (MTN) and from the baseline U-Net architecture (without the regression task) are both assessed against the gold-standard segmentation provided with the dataset as part of the challenge, using four traditionally employed segmentation metrics—Dice Index, Jaccard Index,
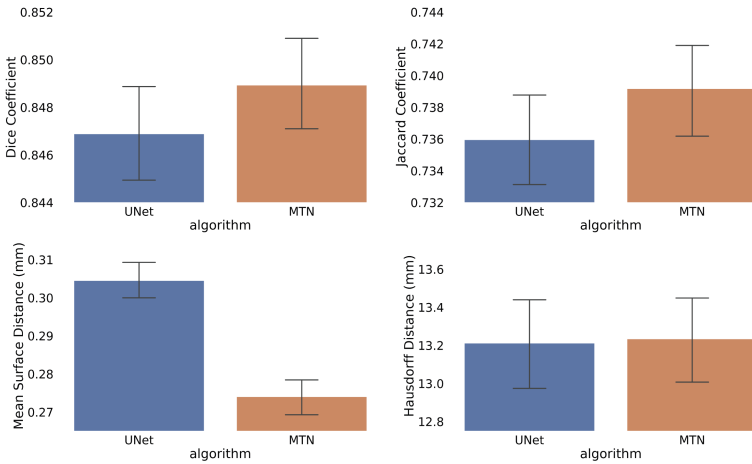

Fig. 2. Mean and 99% confidence interval for (a) dice coefficient, (b) Jaccard coefficient, (c) Mean surface distance (mm), and (d) Hausdorff distance (mm), for baseline U-Net and the proposed MTN architecture across all cardiac phases. Confidence interval is obtained based on 1000 bootstrap re-sampling with replacement for 2191 test volumes across five-fold cross-validation.
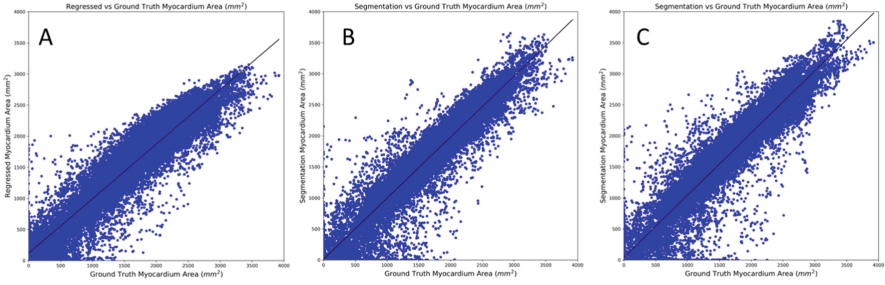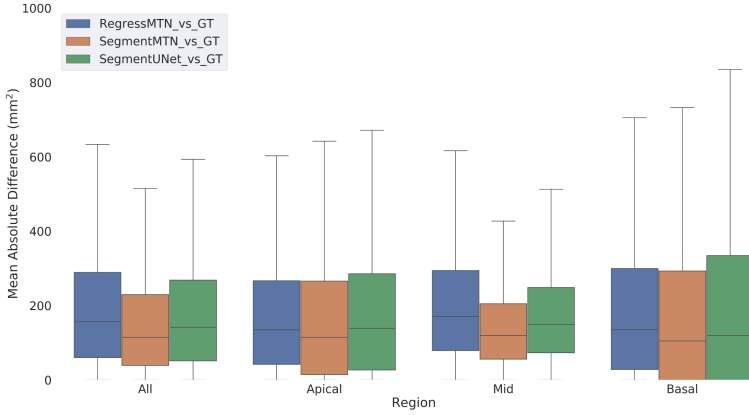
Fig. 3. The myocardium area computed from (A) regression path of the proposed multi-task network, (B) segmentation obtained from the proposed multi-task network, (C) segmentation obtained from the baseline U-Net model, plotted against the corresponding myocardium area obtained from the provided gold-standard segmentation for all cardiac phases. The best fit line is shown in each plot. The correlation coefficients for A, B, and C are 0.9466, 0.9565, 0.9518, respectively

Mean surface distance (MSD), and Hausdorff distance (HD)—summarized in Table 1. Note that the myocardium dice coefficient is higher for end-systole phase where the myocardium is thickest.

The Kolmogorov-Smirnov test shows that the difference in distributions for Dice, Jaccard and MSD metrics between the proposed multi-task network and baseline U-Net architecture are statistically significant with p-values: $2.156e^{-4}$, $2.156e^{-4}$, and $6.950e^{-34}$, respectively. However, since the segmentation is evaluated on a large sample of 2191 volumes across five-fold cross validation, the p-values quickly go to zero even for slight difference in distributions being

Table 2. Mean absolute difference (MAD), in mm^2, between the myocardium area obtained from the provided gold-standard segmentation and the results computed from: (a) the regression path of the proposed multi-task network, (b) segmentation obtained from the proposed multi-task network, and (c) segmentation obtained from the baseline U-Net model, for end-diastole, end-systole, and all cardiac phases, sub-divided into apical, mid, and basal regions of the heart.

Cardiac regions	End-diastole			End-systole			All phases		
	Reg-MTN	Seg-MTN	Seg-UNet	Reg-MTN	Seg-MTN	Seg-UNet	Reg-MTN	Seg-MTN	Seg-UNet
All	201 ± 199	**174** ± 209	203 ± 221	211 ± 209	**173** ± 203	187 ± 204	206 ± 198	**170** ± 199	193 ± 208
Apical	**185** ± 180	187 ± 204	194 ± 186	193 ± 199	190 ± 226	**185** ± 185	184 ± 183	**181** ± 210	187 ± 189
Mid	190 ± 172	**141** ± 132	179 ± 142	228 ± 194	**160** ± 151	174 ± 135	212 ± 178	**149** ± 132	176 ± 141
Basal	**250** ± 269	252 ± 331	282 ± 368	193 ± 241	**186** ± 267	216 ± 312	213 ± 248	**210** ± 289	237 ± 319

(a) Box-plot for the mean absolute difference (MAD).
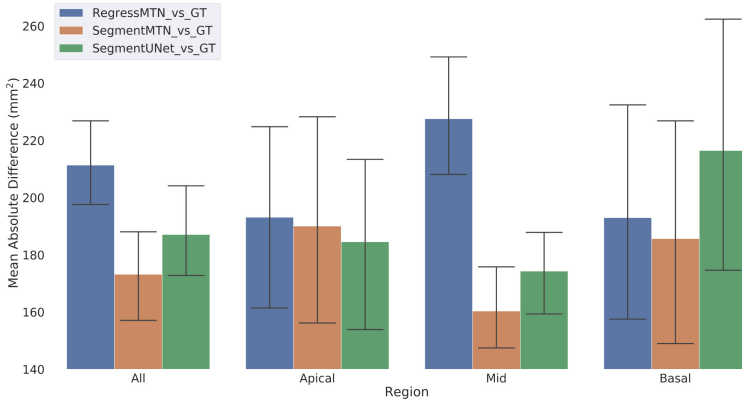
(b) Mean and 99% confidence interval for the mean absolute difference (MAD).

Fig. 4. (a) Box-plot (outliers removed for clarity) and (b) Mean and 99% confidence interval, for the mean absolute difference (MAD) between the myocardium area obtained from the provided gold-standard segmentation and the results obtained from: (1) the regression path of the proposed multi-task network, (2) segmentation obtained from the proposed multi-task network, and (3) segmentation obtained from the baseline U-Net model. Confidence intervals are obtained based on 1000 bootstrap re-sampling with replacement.

compared, representing no practical significance [17]. Hence, we computed the 99% confidence interval for the mean value of each segmentation metric based on 1000 bootstrap re-sampling with replacement, as shown in Fig. 2. As evident from Fig. 2, Dice, Jaccard and HD metrics are statistically similar, whereas the reduction in MSD for the proposed multi-task network compared to the baseline U-Net architecture is statistically significant.

In addition to obtaining the myocardium area from the regression path of the proposed network, it can also be computed indirectly from the obtained myocardium segmentation. Hence, we compute and evaluate the myocardium area estimated from three different sources: (a) regression path of the MTN, (b) segmentation obtained from the MTN, and (c) segmentation obtained from the baseline U-Net model. Figure 3 shows the myocardium area obtained from these three methods for all phases of the cardiac cycle plotted against the ground-truth myocardium area estimated from the gold-standard myocardium segmentation provided as part of the challenge data. We can observe a linear relationship between the computed and gold-standard myocardium areas, and the corresponding correlation coefficients for the methods (a), (b), and (c) are 0.9466, 0.9565, 0.9518, respectively.

Further, we computed the MAD between the ground-truth myocardium area and the area estimated by each of the three methods for end-diastole, end-systole, and all cardiac phases (for 26664 slices across five-fold cross validation). For the regional analysis, slices in the ground-truth segmentation after excluding two apical and two basal slices are considered as mid-slices. Table 2 summarizes the mean and standard deviation for the computed MADs. Box-plots (outliers removed for clarity) comparing the three methods for different regions of the heart throughout the cardiac cycle are shown in Fig. 4a. The MAD in myocardium area estimation of $206 \pm 198\,mm^2$ obtained from the regression output of the proposed method is similar to the results presented in [9]: $223 \pm 193\,mm^2$, while acknowledging the limitation that the study in [9] was conducted on a different dataset than our study. Moreover, while the regression output of the proposed network yields good estimate of the myocardial area, the box-plot in Fig. 4a suggests that even further improved myocardial area estimates can be obtained from a segmentation based method, provided that the quality of the segmentation is good.

Lastly, we computed the 99% confidence interval for the mean value of myocardium area MAD based on 1000 bootstrap re-sampling with replacement, as shown in Fig. 4b. This confirms that the myocardium area estimated from the segmentation output of the proposed multi-task network is significantly better than that obtained from the regression output, however, there is no statistical significance between other methods. Furthermore, we can observe the variability in MAD is highest for the basal slices, followed by apical and mid slices.

4   Discussion, Conclusion, and Future Work

We presented a multi-task learning approach to simultaneously segment and quantify myocardial area. We adapt the U-Net architecture, highly successful in medical image segmentation, to perform an additional regression task. The best location to incorporate the regression path into the network is a hyper-parameter, tuned empirically. We found that adding the regression path in the bottleneck or intermediate decoder layers is detrimental for the segmentation performance of the network, likely due to high influence of the skip connections in the U-Net architecture.

Myocardium area estimates obtained from the regression path of the proposed network are similar to the direct estimation-based results found in the literature. However, our experiments suggest that segmentation-based myocardium area estimation is superior to that obtained from a direct estimation-based method. Lastly, the myocardium segmentation obtained from our method is at least as good as the segmentation obtained from the baseline U-Net model.

To test the generalization performance of the proposed multi-task network, we plan to evaluate the network performance using a lower number of training images. Similarly, we plan to extend this work to segment left ventricle myocardium, blood-pool, and right ventricle, and regress their corresponding areas using the Automated Cardiac Diagnosis Challenge (ACDC)[3] 2017 dataset.

References

1. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. **15**(2), 169–184 (2011)
2. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
3. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Ann. Rev. Biomed. Eng. **19**, 221–248 (2017)
4. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: The IEEE Conference on CVPR, June 2015
6. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Rodríguez, J.G.: A review on deep learning techniques applied to semantic segmentation. CoRR abs/1704.06857 (2017)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015)

---

[3] https://www.creatis.insa-lyon.fr/Challenge/acdc/.

8. Zhen, X., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct and simultaneous four-chamber volume estimation by multi-output regression. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 669–676. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_82

9. Xue, W., Islam, A., Bhaduri, M., Li, S.: Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Trans. Med. Imaging **36**(10), 2057–2067 (2017)

10. Xue, W., Brahm, G., Pandey, S., Leung, S., Li, S.: Full left ventricle quantification via deep multitask relationships learning. Med. Image Anal. **43**, 54–65 (2018)

11. Fonseca, C.G., et al.: The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics **27**(16), 2288–2295 (2011)

12. Suinesiaputra, A., et al.: A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. Med. Image Anal. **18**(1), 50–62 (2014)

13. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., Beare, R.: SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. J. Digit. Imaging **31**, 290–303 (2017)

14. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems 30, pp. 5574–5584. Curran Associates, Inc. (2017)

15. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. CoRR abs/1705.07115 (2017)

16. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)

17. Lin, M., Lucas Jr., H.C., Shmueli, G.: Research commentary - too big to fail: large samples and the p-value problem. Inf. Syst. Res. **24**(4), 906–917 (2013)