A Multi-Task Cross-Task Learning Architecture for Ad Hoc Uncertainty Estimation in 3D Cardiac MRI Image Segmentation

S M Kamrul Hasan[1], **Cristian A Linte**[1,2]

[1] Chester F. Carlson Center for Imaging Science, [2] Department of Biomedical Engineering
Rochester Institute of Technology, Rochester, NY

Abstract

Medical image segmentation has significantly benefitted thanks to deep learning architectures. Furthermore, semi-supervised learning (SSL) has recently been a growing trend for improving a model's overall performance by leveraging abundant unlabeled data. Moreover, learning multiple tasks within the same model further improves model generalizability. To generate smooth and accurate segmentation masks from 3D cardiac MR images, we present a Multi-task Cross-task learning consistency approach to enforce the correlation between the pixel-level (segmentation) and the geometric-level (distance map) tasks. Our extensive experimentation with varied quantities of labeled data in the training sets justifies the effectiveness of our model for the segmentation of the left atrial cavity from Gadolinium-enhanced magnetic resonance (GE-MR) images. With the incorporation of uncertainty estimates to detect failures in the segmentation masks generated by CNNs, our study further showcases the potential of our model to flag low-quality segmentation from a given model.

1. Introduction

While deep learning has shown its potential in a variety of medical image analysis problems including segmentation [1], motion estimation [2] etc., many of these successes are achieved at the cost of a large pool of labeled datasets. Obtaining labeled images however is laborious as well as costly, making the adoption of large-scale deep learning models in clinical settings difficult. To address the limited labeled data problem, semi-supervised learning (SSL) [3] has been a growing trend for improving the deep learning model performance through utilizing the unlabeled data. Furthermore, multi-task learning (MTL) [4] techniques have shown promising results for improving the generalizability of any models by jointly tackling multiple tasks through shared representation learning [5]. To date, a number of approaches address SSL along with MTL-based segmentation from MRI including adversarial learning-based method [6], mutual learning-based approach [7] and techniques based on signed distance map [8]. Recent approaches involve integrating uncertainty map into a mean-teacher framework to guide student network [9] for left atrium segmentation. However, this method lacks the geometric shape of semantic objects, leading to poor segmentation at the edges. Li *et al.* [10] proposed an adversarial-based decoder to enforce the consistency between the model predictions on the original data and the data perturbed by adding noise into it.

As a departure from the existing SSL and MTL models, we propose a novel semi-supervised module exploiting adversarial learning and task-based consistency regularization for jointly learning multiple tasks in a single backbone module – uncertainty estimation, geometric shape generation, and cardiac anatomical structure segmentation. The

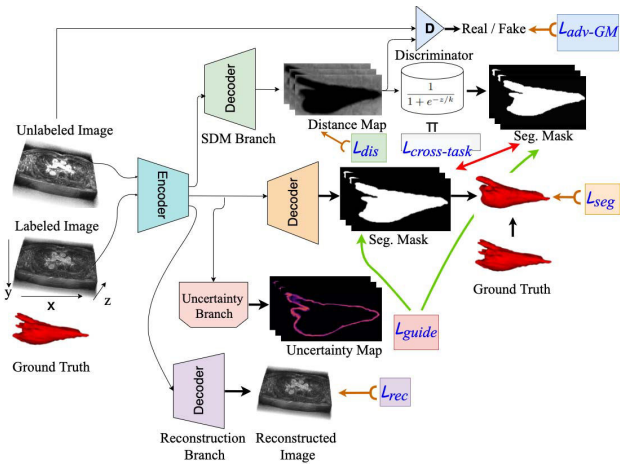

Figure 1: Schematic of the *MTCTL* model: we combine four different decoders who share the same backbone encoder – V-Net.

Corresponding author: S. M. Kamrul Hasan (Email: sh3190@rit.edu). Code, pretrained models, and additional details are available at https://github.com/smkamrulhasan/MTCTL.

2021 Computing in Cardiology (CinC) | 978-1-6654-7916-5/21/$31.00 ©2021110.23919/CINC53138.2021.9662869

network takes as input a 3D volume and outputs an uncertainty map, a 3D distance map, and a segmentation map. The distance map is fed to a transformer to produce a segmentation map which is then used to share the supervisory signal from the predicted segmentation map. To leverage the unlabeled data, the distance map is fed to an adversarial discriminator network to distinguish the predicted distance map from the labeled data. The same encoder backbone is used to estimate the uncertainty of the predicted segmentation map with Monte Carlo sampling.

2. Multi-Task Cross-Task Learning

As shown in Figure 1, our proposed MTCTL model has two distinctive features. First, we combine four different decoders who share the same backbone encoder – V-Net [11]. The uncertainty map generated by the uncertainty decoder is used as the local guidance between the predicted segmentation mask and the mask generated by transforming the distance map. Second, we enforce the correlation between the pixel-level (segmentation) and the geometric-level (distance map) tasks for the generation of smoother and more accurate segmentation masks by introducing the cross-task loss function and include a guidance loss as an uncertainty estimation to smooth out the predicted segmentation mask.

We define the learning task as follows: given an (unknown) data distribution $p(x, y)$ over images and segmentation masks, we have a source domain having a training set, $\mathcal{D}_{\mathcal{L}} = \{(x_1^l, y_1), ..., (x_n^l, y_n\}$ with n labeled data and another domain having a training set, $\mathcal{D}_{\mathcal{UL}} = \{x_1^{ul}, ..., x_m^{ul}\}$ with m unlabeled data which are sampled i.i.d. from $p(x, y)$ and $p(x)$ distribution. Empirically, we want to minimize the target risk $\in_t (\phi, \theta) = min_{\phi,\theta} \mathcal{L}_{\mathcal{L}}(\mathcal{D}_{\mathcal{L}}, (\phi, \theta)) + \gamma \mathcal{L}_{\mathcal{UL}}(\mathcal{D}_{\mathcal{UL}}, (\phi, \theta))$, where $\mathcal{L}_{\mathcal{L}}$ is the supervised loss for segmentation, $\mathcal{L}_{\mathcal{UL}}$ is unsupervised loss defined on unlabeled images and ϕ, θ denotes the learnable parameters of the overall network.

In this work, our architecture is composed of a shared encoder e and a main decoder d, which constitute the segmentation network $f = d \circ e$. We introduce a set of J auxiliary decoders d_a^j, with $j \in [1, J]$.

Dice Loss: For a labeled set $\mathcal{D}_{\mathcal{L}}$, the segmentation network is trained in a traditional supervised manner comprising dice loss, $L_{(seg)}^{\mathcal{L}}(x, y) = \sum_{x_i, y_i \in \mathcal{D}_{\mathcal{L}}} \mathcal{L}_{dice}(x_i, y_i) = \sum_{x_i, y_i \in \mathcal{D}_{\mathcal{L}}} \left[1 - \frac{2 \sum_{x_j \in x_i, y_j \in y_i} f_1(x_i) y_i}{\sum_{x_j \in x_i, y_j \in y_i} f_1(x_j) + \sum_{y_j \in y_i} y_j}\right]$. Then we define the supervised loss for distance map generation task as the mean squared error (MSE) loss between the predicted probability map $f_2(x)$ and the transformed ground truth map $\pi(y)$: $L_{(dis)}^{\mathcal{L}}(x, y) = \sum_{x_i, y_i \in \mathcal{D}_{\mathcal{L}}} ||f_2(x_i) - \pi(y_i)||$.

Smoothing Loss: We utilize a smoothing loss function $L_{(cross-task)}$ to enforce smoothness between the predicted segmentation mask and the inverse transform of the distance map as in [12]: $L_{(cross-task)}(x) = \sum_{x_i \in \mathcal{D}} ||f_1(x_i) - \pi^{-1}(f_2(x_i))||^2 = \sum_{x_i \in \mathcal{D}} ||f_1(x_i) - \frac{1}{1+e^{-k.(f_2(x_i))}}||^2$.

Guidance Loss: As the uncertainty maps give the model some amount of interpretability with which we can decide whether the final segmentation is to be trusted, we consider using Monte-Carlo dropout (MC-dropout) [13] thanks to straightforward implementation. Voxel-wise segmentation uncertainty from MC dropout models is estimated as the mean entropy over all N samples generated by running inference on an input volume N times providing outputs with a set of probability vector of softmax scores, $\{P_n\}_{n=1}^N$ which captures a combination of aleatoric and epistemic uncertainty as, $U(x) = -\frac{1}{N} \sum_{i=1}^N p(x) log(p_i(x))$. We exploit the uncertainty as the guidance to filter out the high uncertainty (unreliable) predictions to minimize the voxel-level mean squared error (MSE) loss between the predicted mask and the transformed mask generated from the distance map: $\mathcal{L}_{\mathcal{G}} = \frac{\sum_{x_i \in (h \times w \times d)} \hat{\mathcal{B}}(U(x) < t) ||f_1(x_i) - \pi^{-1}(f_2(x_i))||^2}{\sum_{x_i \in (h \times w \times d)} \hat{\mathcal{B}}(U(x) < t)}$; where $\hat{\mathcal{B}}(.)$ represents the indicator function for the uncertainty $U(x)$ with threshold t; $f_1(x)$ and $\pi^{-1}(f_2(x_i))$ are the prediction of main decoder and the distance map auxiliary decoder respectively.

Adversarial-Geman-McClure Loss: On the other hand, the data with no corresponding segmentations are trained by minimising the unsupervised loss via a KL divergence which is based on LeastSquares-GAN. However, least-square loss is not robust. Instead, we adopt a new divergence loss function by incorporating it into a Geman-McClure model fashion called *adversarial-Geman-McClure (adv-GM)* loss between the labeled data x_l and the unlabeled data x_{ul}:

$$L_{(adv-GM)}^{\mathcal{U}} = \frac{D\{x^l, dist_l; \phi\}^2 + \{D(x^{ul}, dist_{ul}; \phi) - 1\}^2}{2\beta + D\{x^l, dist_l; \phi\}^2 + \{D(x^{ul}, dist_{ul}; \phi) - 1\}^2}; \quad (1)$$

where $dist_{ul} = f_{dis}(x^{ul}; \theta)$, β is the scale factor which varies in the range of $[0, 1]$ and we set $\beta = 0.5$ in our experiment.

Data: The model was trained and tested on the MICCAI STACOM 2018 Atrial Segmentation Challenge datasets featuring 100 3D gadolinium-enhanced MR imaging scans (GE-MRIs) and LA segmentation masks, with an isotropic resolution of $0.625 \times 0.625 \times 0.625 mm^3$. The dimensions of the MR images may vary depending on each patient, however, all MR images contain exactly 88 slices in the z axis. All the images were normalized and resized to $112 \times 112 \times 80$ before feeding them to the models. We split them into 80 scans for training and 20 scans for validation, and apply the same pre-processing methods.

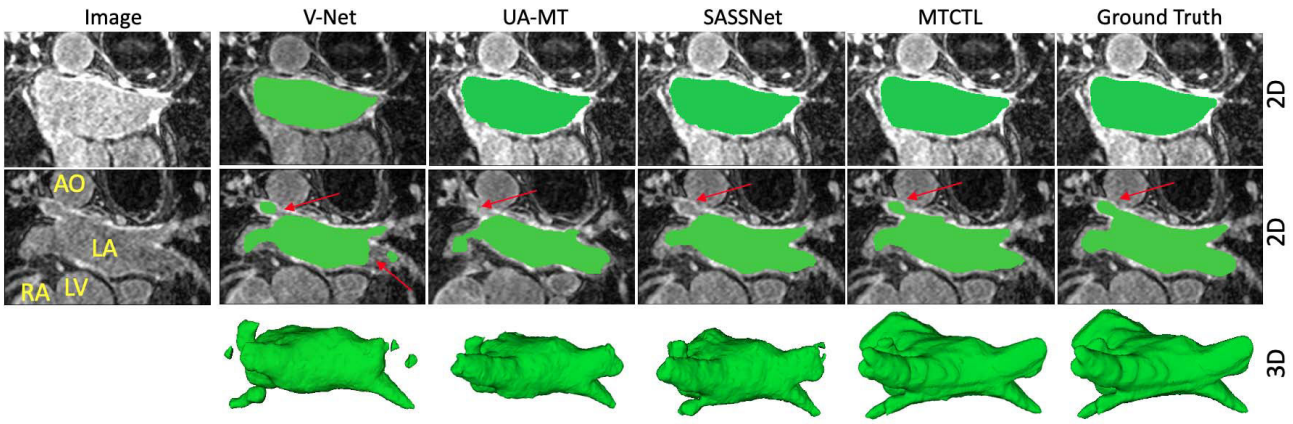

Figure 2: Qualitative comparison of left atrium segmentation result in 2D as well as 3D of the MICCAI STACOM 2018 Atrial Segmentation challenge dataset yielded by four different frameworks: V-Net, UA-MT, SASSNet, and MTCTL. The comparison of segmentation results between the proposed method and three typical deep learning networks indicates that the performance of our proposed network is superior. Red arrow indicates the networks fail to capture the masks near Aorta (AO) region in 3D.

3. Results and Discussion

Figure 2 shows the results obtained by V-Net [11], UA-MT [9], SASSNet [10], our MTCTL, and the corresponding ground truth on the MICCAI STACOM 2018 Atrial Segmentation Challenge from left to right. The second row of the figure shows that all the three frameworks shows a portion of missing masks (red arrow) near Aorta (AO) region, whereas MTCTL generates more complete left atrium segmentation following the addition of multiple tasks (distance map, cross-tasks, and uncertainty guidance) as multiple decoders in either 3D or 2D view.

We conducted a paired statistical test to compare the segmentation performance in Table 1 which shows that our proposed model significantly improved the segmentation performance compared to the semi-supervised, fully-supervised, singletask, and multitask models in terms of the Dice, Jaccard, 95% Hausdorff Distance (95HD), average surface distance (ASD), relative absolute volume difference (RAVD), Precision, and Recall. By exploiting unlabeled data with multiple tasks effectively, our proposed MTCTL model yielded a statistically significant 7.2% improvement ($p < 0.05$) in Dice and 12.5% Jaccard ($p < 0.05$) over the single tasked V-Net framework; a statistically significant 2.5%improvement ($p < 0.05$) in Dice and 4.4% Jaccard ($p < 0.05$) over the SASSNet framework with only 20% labeled training data.

Figure 3 shows a visual comparison of the uncertainty for segmentation of left atrium images in the coronal view. The first and second row presents the uncertainty over segmentation and the uncertainty only for two different slices obtained by the UA-MT and MTCTL framework respectively. In the uncertainty maps, blue pixels have low uncertainty values and red-ish pixels have high uncertainty values. It can be observed from the uncertainty map that the highest uncertainties are located near the border of the segmented foreground, while the pixels with a larger distance to the border have a very low uncertainty.

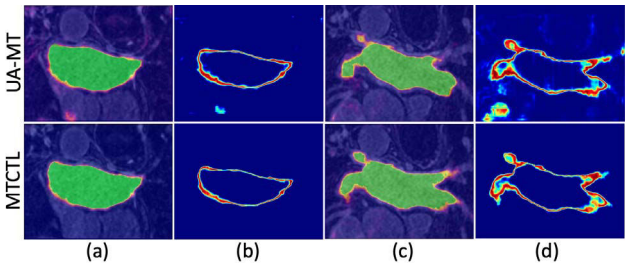

Figure 3: Visual comparison of segmentation predictions overlayed with uncertainty and uncertainty-only (predictive entropy) slices. Segmentation accuracy decreased while predictive uncertainty increased (low uncertainty shown in purple and high uncertainty shown in yellow). Segmentation mask overlaid with uncertainty ((a) & (c)), along with uncertainty maps ((b) & (d)) for two different slices of a patient.

4. Conclusion

In this paper, we proposed a multi-task cross-task learning network (MTCTL) for atrial segmentation. To improve robustness beyond that of the recent SOTA framework, we utilize model uncertainty derived from Monte Carlo Sampling to serve as local guidance between the predicted segmentation mask and the mask generated by transforming

Table 1: Quantitative comparison of left atrium segmentation across several frameworks. Mean (standard deviation) values are reported for $Dice(\%), Jaccard(\%), 95HD(\%), ASD(\%), RAVD(\%), Precision(\%)$, and $Recall(\%)$ from all networks against our proposed MTCTL. The statistical significance of the results for MTCTL model compared against the baseline model SASSNET for 10% and 20% labeled data are represented by $*$ and $**$ for $p-$values 0.1 and 0.05, respectively. The best performance metric is indicated in **bold** text.

	SCANS USED		METRICS						
METHODS	Labeled	Unlabeled	Dice(%) ↑	Jaccard(%)↑	HD95(mm) ↓	ASD(mm)↓	RAVD(%)	Precision(%) ↑	Recall(%)↑
V-Net [11]	10%	0	79.98 ±1.88	68.14±2.01	21.12±15.19	5.47±1.92	-1.34±2.78	83.67±1.79	74.55±1.90
UA-MT [9]	10%	90%	84.25±1.61	73.48±1.73	13.84±13.15	3.36±1.58	-0.13±2.56	87.57±1.53	77.85±1.65
SASSNet [10]	10%	90%	87.32±1.39	77.72±1.49	12.56±11.30	2.55±1.86	-0.09±2.26	87.66±1.38	87.22±1.37
MTCTL (Proposed)	10%	90%	***89.28±0.76**	***80.92±.79**	***7.74±6.05**	**2.0±1.02**	0.56±1.58	***89.74±0.71**	***89.40±0.68**
V-Net [11]	20%	0	85.64±1.73	75.40±1.84	16.96±14.37	4.03±1.53	-0.05±2.64	88.78±1.70	83.79±1.51
UA-MT [9]	20%	80%	88.88±0.73	80.20±0.82	8.13±6.78	2.35±1.16	-2.74±1.58	89.57±0.73	88.82±0.72
SASSNet [10]	20%	80%	89.54±**0.66**	81.24±**0.75**	8.24±6.58	2.27±0.81	0.03±**1.55**	89.86±**0.65**	90.42±**0.66**
MTCTL (Proposed)	20%	80%	****91.80**±0.67	****84.80**±0.83	****5.50**±4.74	**1.55**±0.28	**0.01**±1.65	**91.15**±0.76	**91.04**±0.75

the distance map. Our enforced cross-task loss correlates between the pixel-level (segmentation) and the geometric-level (distance map) tasks to generate smoother and more accurate segmentation masks. We evaluated its performance on the MICCAI STACOM 2018 Atrial Segmentation Challenge dataset. We also conducted an "uncertainty" estimation analysis to determine where our algorithm "fails" to segment regions of interest in an image. Our proposed model outperforms existing methods in terms of both Jaccard and Dice, achieving 89.3% Dice and 80.9% Jaccard with only 10% labeled data and 91.8% Dice and 84.8% Jaccard with only 20% labeled data for atrial segmentation, both of which showed at least 2.5% improvement over the best methods and more than 7% improvement over single-task traditional V-Net architecture.

To our best knowledge, the proposed MTCTL framework constitutes the first approach to adopt adversarial approach along with uncertainty estimation and most accurate semi-supervised left atrium segmentation performance on the LA database. As part of future work, we will use these uncertainty maps to detect regions where the segmentation of the left and right ventricle myocardium and blood pool fails, which is a critical feature for both research and clinical applications.

Acknowledgments

This work was supported by grants from the National Science Foundation (Award No. OAC 1808530) and the National Institutes of Health (Award No. R35GM128877).

References

[1] Hasan SMK, Linte CA. L-CO-Net: Learned condensation-optimization network for segmentation and clinical parameter estimation from cardiac cine MRI. In Int'l Conf. of the IEEE Eng. in Medicine & Biology Society. IEEE, 2020; 1217–1220.

[2] Zheng Q, Delingette H, Ayache N. Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. Medical Image Analysis 2019;56:80–95.

[3] Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020; 12674–12684.

[4] Caruana R. Multitask learning. Machine Learning 1997; 28(1):41–75.

[5] Zhang Y, Wei Y, Yang Q. Learning to multitask. arXiv preprint arXiv180507541 2018;.

[6] Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial learning for semi-supervised semantic segmentation. arXiv preprint arXiv180207934 2018;.

[7] Zhang Y, Zhang J. Dual-task mutual learning for semi-supervised medical image segmentation. arXiv preprint arXiv210304708 2021;.

[8] Dangi S, Linte CA, Yaniv Z. A distance map regularized cnn for cardiac cine MR image segmentation. Medical Physics 2019;46(12):5637–5651.

[9] Yu L, Wang S, Li X, Fu CW, Heng PA. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019; 605–613.

[10] Li S, Zhang C, He X. Shape-aware semi-supervised 3D semantic segmentation for medical images. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2020; 552–561.

[11] Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016; 565–571.

[12] Luo X, Chen J, Song T, Wang G. Semi-supervised medical image segmentation through dual-task consistency. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35. 2021; 8801–8809.

[13] Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? arXiv preprint arXiv170304977 2017;.

Address for correspondence:

S M Kamrul Hasan. RIT – Institute Hall (73) Rm 3130, Rochester NY 14623 USA. E-mail: sh3190@rit.edu