Context-aware Multitasking for Medical Image Segmentation

Aaron Moseley[a], Adam Wang[b], and Abdullah Al Zubaer Imran[a]

[a]University of Kentucky, Lexington, KY, United States
[b]Stanford University, Stanford, CA, United States

ABSTRACT

Fully-supervised medical image segmentation models require a large amount of manually-annotated training data. This expensive, time-consuming, and error-prone annotation process hinders progress and makes it challenging to perform effective segmentations. It is, therefore, imperative that the models learn as efficiently as possible from the limited available data. To this end, we propose a novel multitask learning approach for medical image segmentation models, namely *context-aware multitasking U-Net (CMU-Net)*. This new approach combines the segmentation task with an auxiliary binary classification task focusing on the correct identification of the organs-of-interest in each slice. Such context-awareness provides additional leverage for the main segmentation task in a multitask learning setting. Experimental evaluations on the public LiTS dataset demonstrate superior performance of CMU-Net in segmenting liver from abdominal CT images, achieving an improvement of 4.86% in Dice score and 22.54% in Hausdorff distance.

Keywords: abdomen CT, liver, image segmentation, U-Net, multitask learning

1. DESCRIPTION OF PURPOSE

Medical image segmentation is one of the most important tasks in an imaging pipeline as it influences a number of image-guided decisions. Deep learning models, particularly U-Net, have become the preferred method for the segmentation of medical images.[1] The incorporation of a secondary classification task to improve segmentation results has been a major focus area for previous research. The primary method of accomplishing this is by pre-training segmentation models on non-medical datasets like ImageNet.[2] While this showed improvement over standard U-Net, the differences between medical and natural images imply that the benefits of this method could not be fully exploited in the medical imaging domain. In order to transfer this improvement to the task of medical image segmentation, we incorporate a surrogate classification task into the segmentation training process. We propose an innovative context-aware multitask learning U-Net via training the model on understanding context–presence or absence of organs-of-interest in an input image. Such auxiliary task can be created as a self-supervised pretext[3] or exploiting already available labels. Our presented work takes the latter approach by creating binary classification task from the segmentation labeled dataset. We further investigate different ways the auxiliary classification task could be leveraged: pretraining an U-Net encoder and jointly training the full U-Net model for segmentation.[4] Multitask learning can facilitates optimizing multiple tasks within the same model where one task regularizes another.[5] Training the model on learning the auxiliary classification tasks helps improve the generalizability of the model in a multitask learning setting.[4]

2. METHODS

To formulate the problem, we assume a data distribution $p(X, Y)$ over $\mathcal{D}$ where X is a set of abdominal CT slices and Y is the set of corresponding segmentation maps. Our goal is to create a network G_ϕ with parameters ϕ such that $G_\phi(X) \rightarrow Y$. We also manufacture the secondary classification task by creating a set of class labels C such that $C = \{\max(Y_i)\}$ where $Y_i = \{0, 1\}^{256 \times 256}$ for $i = 1, ..., |Y|$.

To train the CMU-Net network, we use the loss function $L = w_s * d(\hat{y}_s, y_s) + w_c * f(\hat{y}_c, y_c)$ where w_s and w_c are the segmentation and classification weights respectively and $d(\hat{y}_s, y_s)$ and $f(\hat{y}_c, y_c)$ are the segmentation and classification loss functions respectively. Based on a preliminary investigation, we selected weights of $w_s = 0.5$ and $w_c = 0.5$. This training process is shown in Fig. 1.

Send correspondence to A. Imran; E-mail: aimran@uky.edu

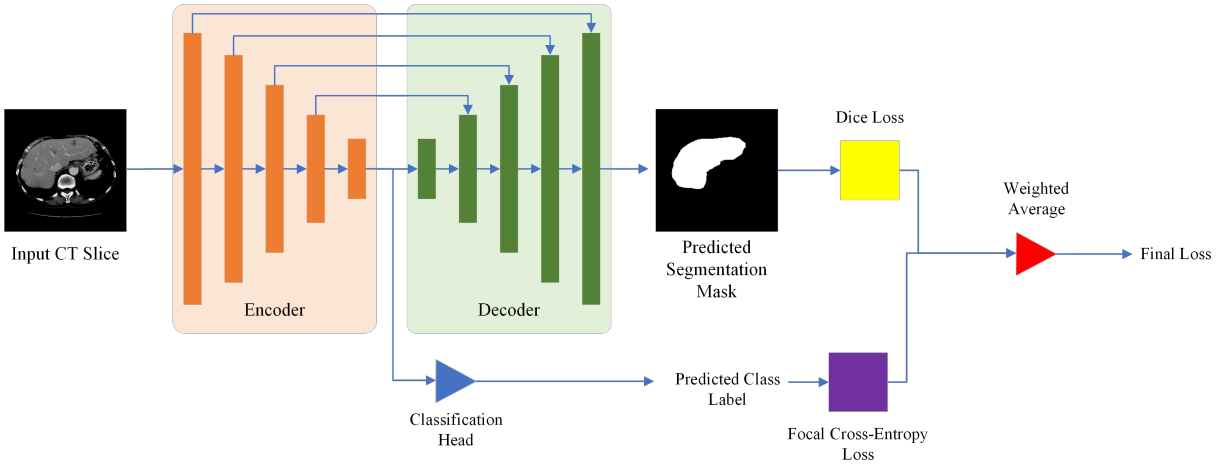

Figure 1. Illustration of the proposed CMU-Net for liver segmentation from abdominal CT: When training, the U-Net encoder outputs a class label prediction while the entire model outputs a predicted segmentation map. The loss values are calculated individually for both outputs, then combined in a weighted average to find the final loss value.

When training the pre-trained U-Net, we followed a similar process but separated the training into two phases. During the first phase, we trained the encoder on binary classification. After training the encoder, we added a decoder to complete the U-Net network and trained the entire network on the binary segmentation task.

For segmentation loss, we used standard dice loss. For classification loss, we used focal cross-entropy loss.[6] This was shown in our preliminary trials to improve the performance of the encoder over standard cross-entropy loss. Our networks are based on the original U-Net architecture but incorporate some of the improvements found in.[4] The encoder section is made of 5 convolutional blocks. Each block consists of two sets of convolutional layers, Leaky ReLU (slope=0.2), instance normalization, and max pooling. For the classification task, a head is added to the encoder that consists of global average pooling and a single fully-connected layer. The decoder also consists of 5 convolutional blocks, each followed by up-sampling and concatenation with a skip connection from the encoder. The encoder begins by using 16 channels and doubles the number of channels after every convolution block, ending at 256. This is mirrored by the decoder, beginning at 256 channels and ending at 16. To generate the segmentation mask, the decoder then uses a 1×1 convolution.

3. EXPERIMENTS AND RESULTS

We trained and evaluated the models using 17 abdominal CT scans from the LiTS dataset.[7] We split the scans into train (10), validation (2), and test (5) sets. The models were trained by extracting 1,664 2D axial slices from the train set, the version of each model was selected based on its performance in each epoch on the 298 slices extracted from the validation set, and each model was evaluated on 787 slices extracted from the test set. For the secondary classification task, we created binary image-level labels (0 for non-liver and 1 for liver) for each slice from their segmentation masks. Each model, including each pre-trained encoder, was trained for 100 epochs with cosine-annealing learn rate scheduling with warm restarts every 10 epochs. After preliminary trials, we found that baseline U-Net and pre-trained U-Net both performed best when using a batch size of 10 and a learn rate of 0.01. CMU-Net performed best when using a batch size of 6 and a learn rate of 0.001. We also found that using focal loss weights of 0.5 and 0.5 for positive and negative examples respectively worked best for pre-trained encoders while weights of 0.8 and 0.2 worked best for CMU-Net. All the models are trained using the same architecture and only differ in their training strategies. Each model was evaluated on both Dice coefficient and Hausdorff distance. Table 1 reports our findings. We found that although there are small differences between the performance of baseline U-Net and pre-trained U-Net, they are not significant. But, we found that the improvement of CMU-Net over standard U-Net is significant across both Dice coefficient ($p < 0.05$) and Hausdorff distance ($p < 0.01$). Both qualitatively (see Fig. 2) and quantitatively, our CMU-Net outperforms the baseline models. This indicates that

Table 1. Performance comparison of our CMU-Net with the baseline U-Net and pretrained U-Net (PU-Net) in segmenting liver from input CT images. We report the average Dice score (DS) and Hausdorff distance (HD) along with the standard deviation, after 10 runs of each model.

Model	Dice Coefficient	Hausdorff Distance
U-Net	0.8567 ± 0.0351	15.7667 ± 2.2025
Pre-Trained U-Net	0.8583 ± 0.0328	15.7389 ± 4.1962
CMU-Net	**0.8984 ± 0.018**	**12.2134 ± 4.2547**

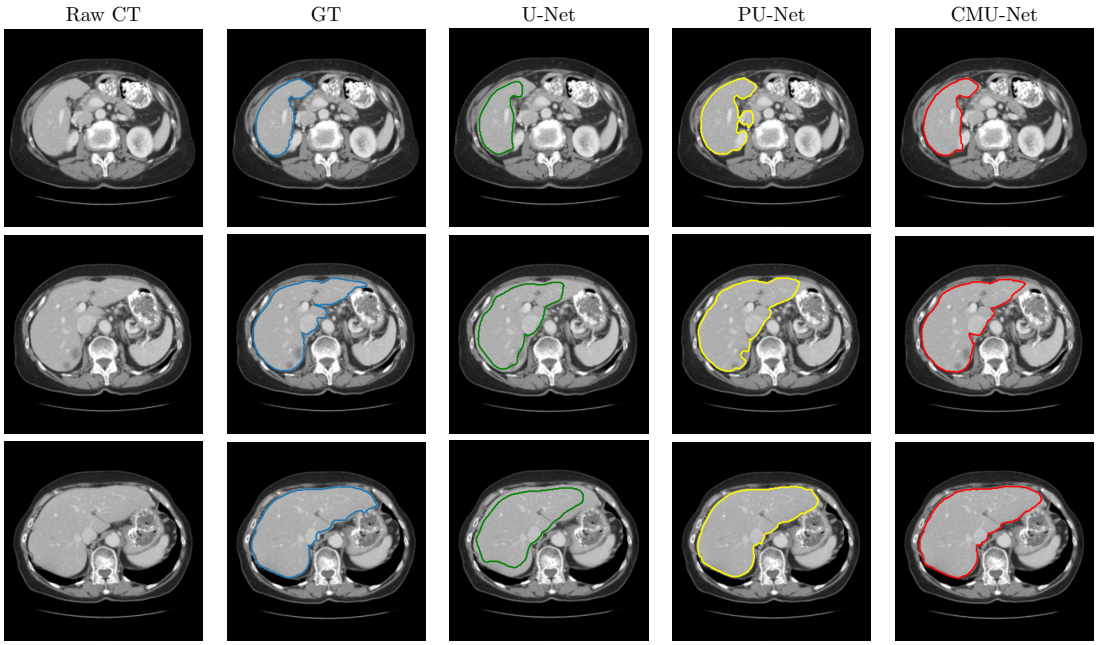

Figure 2. Qualitative comparison of the segmented liver mask by our proposed CMU-Net demonstrates superior results over the baseline models.

leveraging a secondary classification task during training improves the performance of the target segmentation task.

[8]

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

A novel pre-training and joint training approach by creating an innovative surrogate for medical image segmentation validated on segmenting liver from abdominal CT scans.

5. CONCLUSIONS

This work shows that performance improvements can be found through our proposed, previously unexplored, joint training method for segmentation models. We've found that by incorporating a secondary binary classification task for in-domain medical image data, we can significantly improve medical image segmentation models without changing their architecture. Our results are promising, but this topic needs to be explored further to determine whether more benefits can be found. In the future, we plan to expand on the ideas presented here by freezing the U-Net encoder after pre-training, combining pre-training and joint-training into a two-stage process, and incorporating a ResNet backbone into the U-Net encoder.

6. ACKNOWLEDGEMENTS

This work has not been submitted for publication or presentation anywhere else.

REFERENCES

[1] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: convolutional networks for biomedical image segmentation," *CoRR* **abs/1505.04597** (2015).

[2] Iglovikov, V. and Shvets, A., "Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation," *CoRR* **abs/1801.05746** (2018).

[3] Imran, A.-A.-Z., Huang, C., Tang, H., Fan, W., Xiao, Y., Hao, D., Qian, Z., Terzopoulos, D., et al., "Self-supervised, semi-supervised, multi-context learning for the combined classification and segmentation of medical images (student abstract)," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **34**(10), 13815–13816 (2020).

[4] Haque, A., Wang, A., and Terzopoulos, D., "Multimix: sparingly-supervised, extreme multitask learning from medical images," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 693–696, IEEE (2021).

[5] Imran, A.-A.-Z., Pal, D., Patel, B., Wang, A., et al., "SSIQA: multi-task learning for non-reference CT image quality assessment with self-supervised noise level prediction," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 1962–1965, IEEE (2021).

[6] Lin, T., Goyal, P., Girshick, R. B., He, K., and Dollár, P., "Focal loss for dense object detection," *CoRR* **abs/1708.02002** (2017).

[7] Bilic, P., Christ, P., Li, H. B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G. E. H., Chartrand, G., et al., "The liver tumor segmentation benchmark (LiTS)," *Medical Image Analysis* **84**, 102680 (2023).

[8] Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V., and Norouzi, M., "Big self-supervised models advance medical image classification," (2021).