Probabilistic Interactive Segmentation for Medical Images

Hallee E. Wong
MIT CSAIL
hallee@mit.edu

John V. Guttag
MIT CSAIL
guttag@mit.edu

Adrian V. Dalca
MIT CSAIL, HMS/MGH
adalca@mit.edu

Abstract

Deep learning models are effective for medical image analysis tasks such as segmentation. However, training these models requires substantial amounts of labeled data, most often annotated manually. Segmenting new medical images to create labeled training data is a tedious and time-consuming process for human annotators. *Interactive* segmentation tools seek to alleviate this problem, most often by predicting completed segmentations from limited user inputs. This works reasonably well for some domains and for well-defined tasks. But for a new domain or task, the segmentation task is ambiguous. We hypothesize than in such situations proposing multiple partial segmentations is more useful than proposing a single complete segmentation. We propose a *probabilistic partial* segmentation model, that takes an input image and partial segmentation, and predicts possible next steps for the segmentation. The proposed model can be used iteratively to help annotators accurately and efficiently segment new medical images. The user can choose among multiple predicted larger segmentations and perhaps make a small number of corrections before inputting the updated segmentation back into the system. By predicting multiple larger partial segmentations at each iteration rather than attempting to fully complete the segmentation in one step, the system can enable users to produce accurate segmentations for new medical image domains with fewer corrections. We use synthetic data to demonstrate the proposed model and show a proof-of-concept for the system.

1 Introduction

Deep learning models have been successful at performing medical image segmentation [13]. However, training these models requires substantial amounts of labeled data, most often annotated manually. Segmenting new medical images to create labeled training data is a tedious and time-consuming process for human annotators, particularly for 3D modalities. *Interactive* segmentation methods seek to reduce the effort and time required for manual segmentation. Given the diversity of biomedical imaging available, effective tools for interactive segmentation would be most impactful in helping to segment previously unseen regions, and image types where there are little or no existing labelled training data.

Related Works. Most learning-based interactive segmentation systems take a semi-automatic approach, in which the user provides a few initial inputs such as scribbles [15, 16], bounding boxes [6, 9, 11, 15, 17, 18], or clicks [5, 10, 12], and an automatic segmentation model produces a complete prediction that the user corrects and refines. An alternative approach is to use user inputs as a noisy signal to train an automatic segmentation model with weak supervision [9, 10]. However, this approach is not interactive at inference time and requires training using a large dataset, which may not be available for many medical imaging modalities and tasks.

Existing frameworks for interactive segmentation of medical images have focused on minimizing user interaction [5, 10, 12, 18] and developing domain-specific methods with limited generalizability [1, 4, 8, 9]. Deep learning-based techniques [5, 12, 15, 16] have gained popularity because of their ability to

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

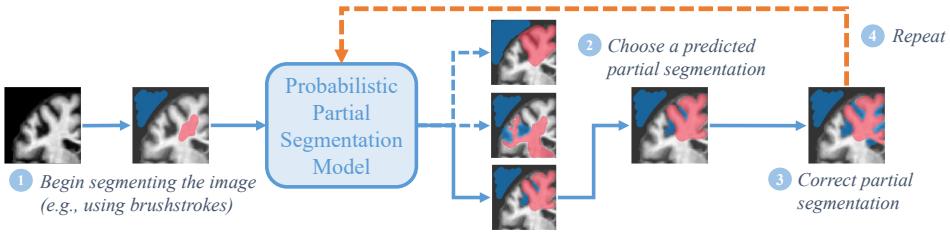

Figure 1: Overview. The user begins segmenting the image (e.g., using brushstrokes) to indicate interior and exterior regions. During each iteration, the proposed model predicts multiple partial segmentations that the user chooses among. The user can then update the segmentation to make corrections.

learn high-level semantic features. In these systems, an automatic segmentation model is used to predict a complete segmentation from user inputs, and then different strategies are used to incorporate additional user corrections. To improve the accuracy of the initial segmentation, some works employ a separate refinement network [5, 16], or learn image-specific models for fine-tuning [15]. A few methods can generalize to new classes and modalities, but the results are limited to a few relatively similar tasks [5, 12, 15].

Contributions. To develop an interactive segmentation system that can be used on a new domain or segmentation target, we propose a *probabilistic partial* segmentation model that i) enlarges a partial segmentation rather than completing it in one step, and ii) produces probabilistic predictions, to give the annotator flexibility. Predicting larger segmentations instead of the whole segmentation enables generalization to new medical segmentation tasks. We focus on predicting several possible segmentations, to enable the user to choose the next step in ambiguous situations and reduce the number of corrections they have to make.

At each iteration the proposed model takes in an input image and the partial segmentation completed so far, and probabilistically predicts a set of possible next steps for the segmentation i.e., larger partial segmentations (Figure 1). The user can then choose a partial segmentation with which to continue. At the end of each iteration, the user can make corrections to the partial segmentation before the model is run again to produce new predictions.

2 Methods

Given an image and initial partial segmentation, we propose a probabilistic model that can predict larger partial segmentations. Given an input $x^{(i)}$ containing an image and partial segmentation, we assume a partial segmentation $y^{(i)} = \{y_m^{(i)}, y_s^{(i)}\}$ at step i, where $y_m^{(i)}$ is a binary partialness mask indicating the pixels that have been segmented and $y_s^{(i)}$ is a binary segmentation mask. We let $x^{(i)} = \{x_0, x_m^{(i)}, x_s^{(i)}\}$ contain an input image x_0, an input binary partialness mask $x_m^{(i)}$, and an input binary segmentation mask $x_s^{(i)}$. For the input partial segmentation $\{x_m^{(i)}, x_s^{(i)}\}$ and larger partial segmentation $\{y_m^{(i)}, y_s^{(i)}\}$ we assume $p_\theta(x_s^{(i)} = 1 | \, x_m^{(i)} = 0) = 1$, $p_\theta(y_s^{(i)} = 1 | \, y_m^{(i)} = 0) = 1$, and $p_\theta(y_s^{(i)} = x_s^{(i)} | \, x_m^{(i)} = 1) = 1$. We let $z^{(i)}$ be a latent variable that captures the intrinsic variability of the partial segmentation $y^{(i)}$ for input x, and assume a prior $p_\theta(z; x) \sim \mathcal{N}(0, I)$. We let $y^{(i)}$ be a noisy observation of the non-linear decoding g_θ of $z^{(i)}$: $p_\theta(y^{(i)}|z^{(i)}; x^{(i)}) \propto \exp\{-f\big(y^{(i)}, g_\theta(z^{(i)}, x^{(i)}); x^{(i)}\big)\}$ where f is a partial segmentation distance measure.

Give a set of partial segmentations $\{y^{(i)}\}_{i=1}^N$ we aim to maximize the posterior probability $p_\theta(z|y; x)$. Unfortunately, computing the posterior probability $p_\theta(z|y; x)$ is intractable. We therefore follow a variational strategy and introduce an approximate posterior probability $q_\phi(z|y; x)$. We minimize the KL divergence between the real and approximate posterior, which leads to maximizing the evidence lower bound objective (ELBO),

$$\log p_\theta(y|x) \geq \mathbb{E}_{q_\phi(z|y;\, x)}\big[\log p_\theta(y|z; x)\big] - KL\big(q_\phi(z|y; x) \,||\, p_\theta(z; x)\big). \tag{1}$$

Figure 2 shows the proposed setup, using a decoder network to model $p_\theta(y|z; x)$ and an encoder network to model $q_\phi(z|y; x) = \mathcal{N}(\mu_\phi(x^{(i)}, y^{(i)}), \sigma_\phi(x^{(i)}, y^{(i)}))$, where $\mu_\phi(x^{(i)}, y^{(i)})$ and $\sigma_\phi(x^{(i)}, y^{(i)})$ are differentiable deterministic functions similar to a conditional variational autoencoder [3, 14].

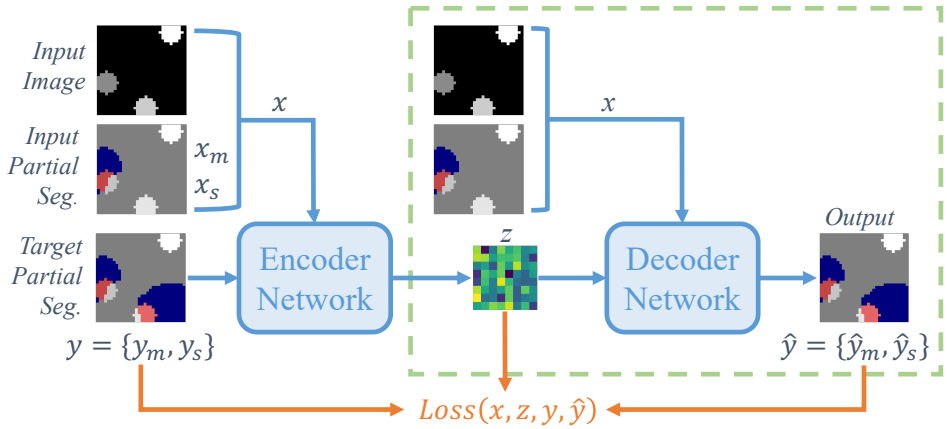

Figure 2: Proposed probabilistic partial segmentation model. Inputs are an image, a partial segmentation and a larger target partial segmentation. At inference (green box) we use a decoder $g_\theta(z; x)$ to predict partial segmentation $\hat{y} = \{\hat{y}_m, \hat{y}_s\}$ given an image and partial segmentation x and a randomly sampled encoding $z \sim \mathcal{N}(0, I)$.

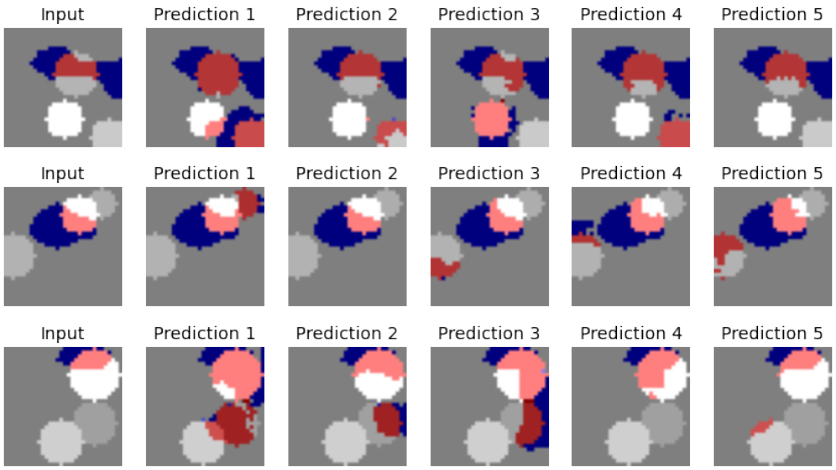

Figure 3: Multiple predicted binary partial segmentations generated using the decoder with different $z \sim N(0, I)$. Segmented interior (red) and exterior (blue) are overlayed on the input image. The decoder produces plausible partial segmentations that segment multiple possible subsets of disks.

We let the partial segmentation distance $f(\cdot,\cdot)$ be

$$f(y, \hat{y}; x) = \text{WDL}(y_s, \hat{y}_s, y_m - x_m) + \text{DL}(y_m, \hat{y}_m) \tag{2}$$

where $\text{DL}(y_m, \hat{y}_m)$ is (soft) Dice loss [7] comparing the target and predicted partialness masks and $\text{WDL}(y_s, \hat{y}_s, y_m - x_m)$ is (soft) Dice loss comparing segmentation masks weighted by the target change in partialness mask $y_m - x_m$. Our final loss

$$\mathcal{L}(\theta, \phi, x, y, z) = \text{WDL}\big(y_s, \hat{y}_{\theta s}, y_m - x_m\big) + \beta_0 \cdot \text{DL}(y_m, \hat{y}_{\theta m}) + \beta_1 \cdot \text{KL}\big(\mathcal{N}(z; \mu_\phi, \sigma^2_\phi), \mathcal{N}(z; 0, I)\big) \tag{3}$$

follows from maximizing the conditional log likelihood using the ELBO.

3 Experiments

Data. We created a preliminary dataset of 1,000 images of disks with corresponding ground truth segmentation maps of different subsets of disks. Each 32×32 image contains three randomly generated disks of different random intensity values. During training we simulate pairs of partial segmentations by taking the intersection of the ground truth segmentation with synthetic images of random shapes [2].

Setup. We train encoder and decoder networks that minimize (3) and set $\beta_0 = 1$ and $\beta_1 = 10^{-3}$. The encoder is a 3-layer CNN with 5, 32, and 16 channels with kernel size 3, ReLU activation, and 2x down-sampling using max pooling after each layer. The decoder is a 4-layer CNN with an initial up-sampling layer to resize z to be the same spatial dimensions of as x, a 2D convolutional layer, 2x down-sampling using max pooling, a 2D convolutional layer, a 2D transpose convolutional layer and a 2D convolutional layer. The convolutional layers have 4, 64, 64, and 64 channels with kernel size 3 and ReLU activation, except for the final learnable layer, which uses sigmoid activation. We enforce that the input partial segmentation $\{x_m, x_s\}$ is preserved in the predicted partial segmentation $\{\hat{y}_m, \hat{y}_s\}$ and the predicted segmentation $\hat{y}_s$ only contains segmentation in areas covered by the partialness mask $\hat{y}_m$ i.e., $p_\theta(\hat{y}_s = 1, \hat{y}_m = 0) = 0$.

Results. Figure 3 shows that the proposed decoder network produces larger partial segmentations that are consistent with the input partial segmentation but segment different subsets of disks. For the first input (top row, Figure 3), sampling different z changes whether the two bottom disks are segmented or not. Sampling $z \sim \mathcal{N}(0, I)$ for 1,000 held-out validation images, the predicted partial segmentations on average add 52 pixels of foreground segmentation and 31 pixels of background segmentation with a mean Dice of 0.87, taking the maximum Dice on the added partial segmentation compared to all possible ground truth segmentations of disks.

Conclusion. We demonstrate a proof-of-concept for a new framework for probabilistic interactive segmentation using synthetic images. We plan to train the proposed model with a combination of simulated and diverse real data, focused on generalizing to new medical imaging modalities and segmentation tasks where the intentions of the human annotator are ambiguous.

4 Societal Impact

Interactive segmentation tools make it easier for humans to annotate medical images for clinical research and to create labelled datasets for model training and evaluation. Lowering the barrier to manual annotation would make it easier to collect diverse medical imaging datasets. The probabilistic nature of the proposed framework gives annotators more flexibility in their annotations compared to previous tools, however they might still be biased towards particular annotations by the systems' predictions. Errors or bias in the annotation of training data could negatively affect the predictions made by models used for clinical decision making and medical research, and therefore deserve special attention and rigorous experiments, which we plan to undertake.

References

[1] A. Atzeni, L. Peter, E. Robinson, E. Blackburn, J. Althonayan, D. C. Alexander, and J. E. Iglesias. Deep active learning for suggestive segmentation of biomedical image stacks via optimisation of Dice scores and traced boundary length. *Medical Image Analysis*, 81, 2022.

[2] M. Hoffmann, B. Billot, D. N. Greve, J. E. Iglesias, B. Fischl, and A. V. Dalca. SynthMorph: Learning Contrast-Invariant Registration Without Acquired Images. *IEEE Transactions on Medical Imaging*, 41(3):543–558, 2022.

[3] D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations (ICLR)*, 2014.

[4] Q. Liu, Z. Xu, Y. Jiao, and M. Niethammer. iSegFormer: Interactive Segmentation via Transformers with Application to 3D Knee MR Images. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2022.

[5] X. Luo, G. Wang, T. Song, J. Zhang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang. MIDeepSeg: Minimally Interactive Segmentation of Unseen Objects from Medical Images Using Deep Learning. *Medical Image Analysis*, 72:102102, 2021.

[6] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool. Deep Extreme Cut: From Extreme Points to Object Segmentation, 2018. arXiv:1711.09081.

[7] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *Fourth International Conference on 3D Vision (3DV)*, pages 565–571, 2016.

[8] A. Z. H. Ooi, Z. Embong, A. I. Abd Hamid, R. Zainon, S. L. Wang, T. F. Ng, R. A. Hamzah, S. S. Teoh, and H. Ibrahim. Interactive Blood Vessel Segmentation from Retinal Fundus Image Based on Canny Edge Detector. *Sensors*, 21(19):6380, Sept. 2021.

[9] M. Rajchl, M. C. H. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz, and D. Rueckert. DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks. *IEEE Transactions on Medical Imaging*, 36(2):674–683, 2017.

[10] H. R. Roth, D. Yang, Z. Xu, X. Wang, and D. Xu. Going to Extremes: Weakly Supervised Medical Image Segmentation, 2020. arXiv:2009.11988.

[11] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut" — Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics*, 23:309–314, 2004.

[12] T. Sakinis, F. Milletari, H. Roth, P. Korfiatis, P. Kostandy, K. Philbrick, Z. Akkus, Z. Xu, D. Xu, and B. J. Erickson. Interactive segmentation of medical images through fully convolutional neural networks, 2019. arXiv:1903.08205.

[13] D. Shen, G. Wu, and H.-I. Suk. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19(1):221–248, 2017.

[14] K. Sohn, H. Lee, and X. Yan. Learning Structured Output Representation using Deep Conditional Generative Models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28, 2015.

[15] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, 2018.

[16] G. Wang, M. A. Zuluaga, W. Li, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1559–1572, 2019.

[17] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang. Deep GrabCut for Object Selection. arXiv, 2017. arXiv:1707.00243.

[18] S. Zhang, J. H. Liew, Y. Wei, S. Wei, and Y. Zhao. Interactive Object Segmentation With Inside-Outside Guidance. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12231–12241. IEEE, 2020.