

Efficient and Generic Interactive Segmentation Framework to Correct Mispredictions during Clinical Evaluation of Medical Images

Bhavani Sambaturu¹ Ashutosh Gupta² C.V. Jawahar¹ Chetan Arora²

¹ International Institute of Information Technology, Hyderabad, India
`bhavani.sambaturu@research.iiit.ac.in`

² Indian Institute of Technology, Delhi, India

Abstract. Semantic segmentation of medical images is an essential first step in computer-aided diagnosis systems for many applications. However, given many disparate imaging modalities and inherent variations in the patient data, it is difficult to consistently achieve high accuracy using modern deep neural networks (DNNs). This has led researchers to propose interactive image segmentation techniques where a medical expert can interactively correct the output of a DNN to the desired accuracy. However, these techniques often need separate training data with the associated human interactions, and do not generalize to various diseases, and types of medical images. In this paper, we suggest a novel conditional inference technique for DNNs which takes the intervention by a medical expert as test time constraints and performs inference conditioned upon these constraints. Our technique is generic can be used for medical images from any modality. Unlike other methods, our approach can correct multiple structures simultaneously and add structures missed at initial segmentation. We report an improvement of 13.3, 12.5, 17.8, 10.2, and 12.4 times in user annotation time than full human annotation for the nucleus, multiple cells, liver and tumor, organ, and brain segmentation respectively. We report a time saving of 2.8, 3.0, 1.9, 4.4, and 8.6 fold compared to other interactive segmentation techniques. Our method can be useful to clinicians for diagnosis and post-surgical follow-up with minimal intervention from the medical expert. The source-code and the detailed results are available here [1].

Keywords: Machine Learning · Segmentation · Human-in-the-Loop

1 Introduction

Motivation: Image segmentation is a vital imaging processing technique to extract the region of interest (ROI) for medical diagnosis, modeling, and intervention tasks. It is especially important for tasks such as the volumetric estimation of structures such as tumors which is important both for diagnosis and post-surgical follow-up. A major challenge in medical image segmentation is the high variability in capturing protocols and modalities like X-ray, CT, MRI, microscopy, PET, SPECT, Endoscopy and OCT. Even within a single modality, the

Table 1: Comparative strengths of various interactive segmentation techniques.

| Capability | Description | [2] | [3] | [4] | [5] | Ours |
|------------------|-------------------------------------|-----|-----|-----|-----|------|
| Feedback mode | Point | ✓ | ✓ | ✗ | ✓ | ✓ |
| | Box | ✗ | ✗ | ✓ | ✗ | ✓ |
| | Scribble | ✗ | ✗ | ✗ | ✗ | ✓ |
| Training | Pre-training with user interaction | ✓ | ✓ | ✓ | ✓ | ✗ |
| Requirement | Can work with any pre-trained DNN | ✓ | ✓ | ✗ | ✗ | ✓ |
| Correction Modes | Correct multiple labels | ✗ | ✗ | ✗ | ✗ | ✓ |
| | Insert missing labels | ✗ | ✗ | ✗ | ✗ | ✓ |
| Generalization | Adapt: Distribution mismatch | ✗ | ✗ | ✗ | ✗ | ✓ |
| | Segment new organs than trained for | ✗ | ✗ | ✗ | ✗ | ✓ |

human anatomy itself has significant variation modes leading to vast observed differences in the corresponding images. Hence, fully automated state-of-the-art methods have not been able to consistently demonstrated desired robustness and accuracy for segmentation in clinical use. This has led researchers to develop techniques for interactive segmentation which can correct the mispredictions during clinical evaluation and make-up for the shortfall.

Current Solutions: Though it is helpful to leverage user interactions to improve the quality of segmentation at test time, this often increases the burden on the user. A good interactive segmentation method should improve the segmentation of the image with the minimum number of user interactions. Various popular interactive segmentation techniques for medical imaging have been proposed in the literature [6–8]. The primary limitation is that it can segment only one structure at a time. This leads to a significant increase in user interactions when a large number of segments are involved. Recent DNN based techniques [2–4] improve this aspect by reducing user interactions. It exploits pre-learned patterns and correlations for correcting the other unannotated errors as well. However, they require vast user interaction data for training the DNN model, which increases cost and restricts generalization to other problems.

Our Contribution: We introduce an interactive segmentation technique using a pre-trained semantic segmentation network, without any additional architectural modifications to accurately segment 2D and 3D medical images with help from a medical expert. Our formulation models user interactions as the additional test time constraints to be met by the predictions of a DNN. The Lagrangian formulation of the optimization problem is solved by the proposed alternate maximization and minimization strategy, implemented through the stochastic gradient descent. This is very similar to the standard back-propagation based training for the DNNs and can readily be implemented. The proposed technique has several advantages: (1) exhibits the capability to correct multiple structures at the same time leading to a significant reduction in the user time. (2) exploits the learnt correlations in a pre-trained deep learning semantic segmentation net-

work so that a little feedback from the expert can correct large mispredictions. (3) requires no joint training with the user inputs to obtain a better segmentation, which is a severe limitation in other methods [2, 3]. (4) add missing labels while segmenting a structure if it was missed in the first iteration or wrongly labeled as some other structure. The multiple types of corrections allow us to correct major mispredictions in relatively fewer iterations. (5) handle distribution mismatches between the training and test sets. This can arise even for the same disease and image modality due to the different machine and image capturing protocols and demographics. (6) for the same image modality, using this technique one can even segment new organs using a DNN trained on some other organ type. Table 1 summarizes the comparative advantages of our approach.

2 Related Work

Conventional Techniques: Interactive segmentation is a well-explored area in computer vision and some notable techniques are based on Graph Cuts [6, 7, 9], Edge or Active Contours [10, 11], Label propagation using Random Walk or other similar style [8, 12], and region-based methods [13, 14]. In these techniques, it is not possible to correct multiple labels together without the user providing the initial seeds and also not possible to insert a missing label.

DNN based Techniques: DNN based techniques use inputs such as clicks [2, 3], scribbles [15], and bounding boxes [4] provided by a user. Other notable techniques include [3, 4, 16–18]. These methods require special pre-training with user-interactions and associated images. This increases the cost of deployment and ties a solution to pre-decided specific problem and architecture.

Interactive Segmentation for Medical Images: Interactive Segmentation based methods, especially for medical image data, have been proposed in [5, 19–22]. The methods either need the user inputs to be provided as an additional channel with the image [21] or need an additional network to process the user input [19]. BIFSeg [20] uses the user inputs at test time with a DNN for interactive segmentation of medical images. However, our method is significantly different in the following manner: (a) DNN - use their own custom neural networks [20]. However, our method can use pre-existing segmentation networks. This allows our method to use newer architectures which may be proposed in the future as well. (b) Optimization - use CRF-based regularization for label correction [20]. We propose a novel restricted Lagrangian-based formulation. This enables us to do a sample specific fine-tuning of the network, and allows our method to do multiple label corrections in a single iteration which is novel. (c) User Inputs - use scribbles and bounding boxes as user inputs [20]. We can correct labels irrespective of the type of user input provided.

3 Proposed Framework

The goal is to design an approximate optimization algorithm that can encode the constraints arising from user-provided inputs in the form of scribbles. A

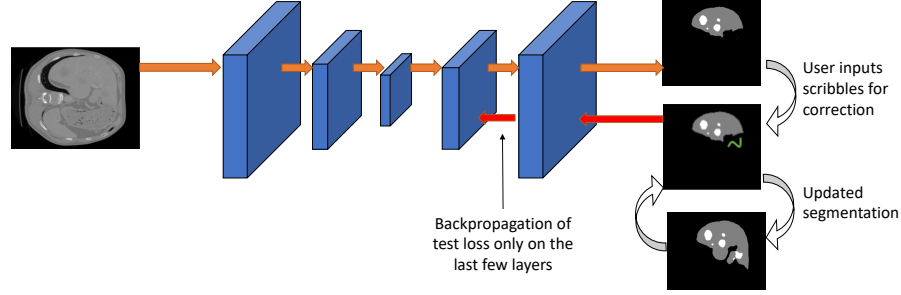


Fig. 1: The figure shows the working of our algorithm. Note that depending upon the application, our framework can use different pre-trained network architectures. Hence we do not give the detailed architecture of any particular model. The first step in our framework is to obtain an initial segmentation using the pre-trained deep learning network. The user then examines the segmentation and adds scribbles where the desired correction is required. This is then used to refine the weights of the network and the improved segmentation is obtained.

simple gradient descent strategy similar in spirit to the Lagrangian relaxation proposed by [23] is optimized. The strategy allows us to use existing libraries and infrastructure built for any image modality optimizing the loss for the DNNs using the standard back-propagation procedure.

Problem Definition: A neural network with N layers is parameterized by weights W from input to output. We represent this as a function $\Psi(x, y, W) \rightarrow \mathbb{R}_+$ to measure the likelihood of a predicted output y given an input x and parameters/weights W . We also want to enforce that the output values belong to a set of scribbles \mathbb{S}^x provided by the user to correct the segmentation dependent on x . Here, \mathbb{S}^x encodes both the location in the image where correction is required and the desired segmentation class label.

We can express the constraint, $y \in \mathbb{S}^x$, as an equality constraint, using a function $g(y, \mathbb{S}^x) \rightarrow \mathbb{R}_+$. This function measures the compatibility between the output y and scribbles \mathbb{S}^x such that $g(y, \mathbb{S}^x) = 0$ if and only if there are no errors in y with respect to \mathbb{S}^x . In our case, $g(y, \mathbb{S}^x)$ is the cross-entropy loss between the predicted labels y and the segmentation class label encoded in \mathbb{S}^x . This allows us to solve the optimization problem by minimizing the following Lagrangian:

$$\min_{\lambda} \max_y \Psi(x, y, W) + \lambda g(y, \mathbb{S}^x). \quad (1)$$

Note that the compatibility constraints in $g(y, \mathbb{S}^x)$ factorize over the pixels and one trivial solution of the optimization problem as described above is to simply change the output variables to the class labels provided by the scribbles. However, this does not allow us to exploit the neighborhood information inherent in the images, and the correlations learnt by a DNN due to prior training over a large dataset.

We note that the network's weights can also control the compatibility of the output configurations with the scribble input. Since the weights are typically

Algorithm 1: Scribble aware inference for neural networks

Input : test instance x , input specific scribbles \mathbb{S}^x , max epochs M ,
pre-trained weights W , η learning rate, α regularization factor
 $W_\lambda \leftarrow W$ // reset to have instance-specific weights

Output : Refined segmentation

while $g(y, \mathbb{S}^x) > 0$ and iteration $< M$ **do**

$y \leftarrow f(x; W_\lambda)$ // perform inference using weights W_λ
 $\nabla \leftarrow g(y, \mathbb{S}^x) \frac{\partial}{\partial W_{\lambda_l}} \Psi(x, y, W_{\lambda_l}) + \alpha \frac{W_l - W_{\lambda_l}}{\|W_l - W_{\lambda_l}\|_2}$ // constraint loss
 $W_{\lambda_l} \leftarrow W_{\lambda_l} - \eta \nabla$ // update instance-specific weights with SGD

end

return y , the refined segmentation

tied across space, the weights are likely to generalize across related outputs in the neighborhood. This fixes the incompatibilities not even pointed-to by the limited scribbles given by the user. Hence, we propose to utilize the constraint violation as a part of the objective function to adjust the model parameters to search for an output satisfying the constraints efficiently.

We propose to optimize a “dual” set of model parameters W_λ over the constraint function while regularizing W_λ to stay close to the original weights W . The network is divided into a final set of layers l and an initial set of layers $N-l$. We propose to optimize only the weights corresponding to the final set of layers W_{λ_l} . The optimization function is given as:

$$\min_{W_{\lambda_l}} \Psi(x, \hat{y}, W_{\lambda_l}) \quad g(\hat{y}, \mathbb{S}^x) + \alpha \|W_l - W_{\lambda_l}\|, \quad (2)$$

where $\hat{y} = \arg \max_y \Psi(x, y, W_{\lambda_l})$. This function is reasonable by definition of the constraint loss $g(\cdot)$, though it deviates from the original optimization problem, and the global minima should correspond to the outputs satisfying the constraints. If we initialize $W_\lambda = W$, we also expect to find the high-probability optima. If there is a constraint violation in \hat{y} , then $g(\cdot) > 0$, and the following gradient descent procedure makes such \hat{y} less likely, else $g(\cdot) = 0$ and the gradient of the energy is zero leaving \hat{y} unchanged.

The proposed algorithm (see Algorithm 1) alternates between maximization to find \hat{y} and minimization w.r.t. W_{λ_l} to optimize the objective. The maximization step can be achieved by employing the neural network’s inference procedure to find the \hat{y} , whereas minimizing the objective w.r.t. W_{λ_l} can be achieved by performing stochastic gradient descent (SGD) given a fixed \hat{y} . We use the above-outlined procedure in an iterative manner (multiple forward, and back-propagation iterations) to align the outcome of the segmentation network with the scribble input provided by the user.

Fig. 1 gives a visual description of our framework. It explains the stochastic gradient-based optimization strategy, executed in a manner similar to the standard back-propagation style of gradient descent. However, the difference is that while the back-propagation updates the weights to minimize the training loss,

the proposed stochastic gradient approach biases the network output towards the constraints generated by the user provided scribbles at the test time.

Scribble Region Growing: The success of an interactive segmentation system is determined by the amount of burden on a user. This burden can be eased by allowing the user to provide fewer, shorter scribbles. However, providing shorter scribbles can potentially entail a greater number of iterations to obtain the final accurate segmentation. Hence, we propose using region growing to increase the area covered by the scribbles. We grow the region to a new neighborhood pixel, if the intensity of the new pixel differs from the current pixel by less than a threshold T .

4 Results and Discussions

Dataset and Evaluation Methodology: To validate and demonstrate our method, we have evaluated our approach on the following publicly available datasets containing images captured in different modalities: **(1) Microscopy:** 2018 Data Science Bowl (2018 DSB) [24] (nucleus), MonuSeg [25] (nucleus), and ConSeP [26] datasets (epithelial, inflammatory, spindle shaped and miscellaneous cell nuclei) **(2) CT:** LiTS [27] (liver and tumor cells) and SegThor [28] (heart, trachea, aorta, esophagus) challenges **(3) MRI:** BraTS' 15 [29] (necrosis, edema, non-enhancing tumor, enhancing tumor) and CHAOS [30] (liver, left kidney, right kidney, spleen) datasets. All the experiments were conducted in a Linux environment on a 20 GB GPU (NVIDIA 2018Tx) on a Core-i10 processor, 64 GB RAM, and the scribbles were provided using the WACOM tablet. For microscopy images, the segmented image was taken and scribbles were provided in areas where correction was required using LabelMe [31]. For CT and MRI scans, the scribbles were provided in the slices of the segmentation scan where correction was desired using 3-D Slicer [32]. For validating on each of the input modalities, and the corresponding dataset, we have taken a recent state-of-the-art approach for which the DNN model is publicly available and converted it into an interactive segmentation model. We used the same set of hyper-parameters that were used for training the pre-trained model. The details of each model, and source code to test them in our framework are available at [1]. To demonstrate the time saved over manual mode, we have segmented the images/scans using LabelMe for microscopy, and 3-D Slicer for CT/MRI, and report it as full human annotation time (FH). We took the help of two trained annotators, two general practitioners and a radiologist for the annotation.

Ablation Studies: We also performed ablation studies to determine : (a) Optimum number of iterations, (b) Layer number upto which we need to update the weights, (c) Type of user input (point,box,scribble) and, (d) Effect of scribble length on the user interaction time. Owing to space constraints, the result of the ablation studies are provided on the project page [1]. We find scribble as the most efficient way of the user input through our ablation study, and use them in the rest of the paper.

Table 2: User Interaction Time (**UT**) and Machine Time (**MT**) in minutes to separate structures (**FH**: Full Human Annotation, **RG**: Our method - Region Growing, **NRG**: Our Method - No Region Growing. Methods [3, 4, 6–8, 33] were applied till a dice coefficient of 0.95 was reached.

| Dataset | User Interaction Time | | | | | | | | | Machine Time | | | | | | | |
|-----------|-----------------------|----|-----|-----|-----|-----|------|-----|-----|--------------|-----|-----|-----|-----|------|-----|-----|
| | FH | RG | NRG | [6] | [3] | [4] | [33] | [7] | [8] | RG | NRG | [6] | [3] | [4] | [33] | [7] | [8] |
| 2018 DSB | 66 | 5 | 7 | 13 | 12 | 12 | - | - | - | 6 | 10 | 11 | 12 | 13 | - | - | - |
| CoNSeP | 30 | 6 | 8 | 16 | 18 | 20 | - | - | - | 5 | 7 | 17 | 20 | 23 | - | - | - |
| LiTS | 120 | 7 | 8 | - | - | - | 11 | 12 | 13 | 10 | 12 | - | - | - | 11 | 13 | 11 |
| CHAOS | 136 | 13 | 15 | - | - | - | 58 | 66 | 83 | 25 | 30 | - | - | - | 50 | 66 | 83 |
| BraTS' 15 | 166 | 11 | 13 | - | - | - | 76 | 83 | 100 | 58 | 81 | - | - | - | 100 | 116 | 133 |

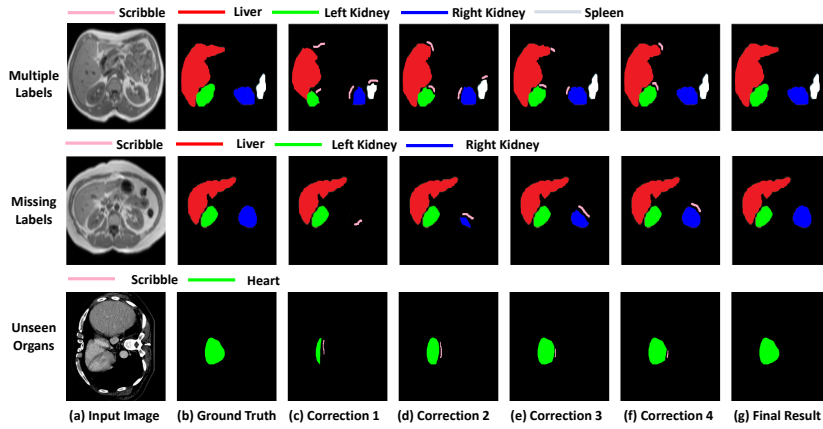


Fig. 2: (a) Correcting multiple labels (top row) (b) Inserting missing labels (middle row) (c) Interactive segmentation of organs the model was not trained for (bottom row). Incremental improvement as scribbles are added shown. No other state-of-the-art approach has these capabilities. More qualitative results are provided here [1].

Image Segmentation with Multiple Classes: Our first experiment is to evaluate interactive segmentation in a multi-class setting. We use two trained annotators for the experiment. We have used the validation sets of the 2018 Data Science Bowl (2018 DSB), CoNSeP, LiTS, CHAOS and the BraTS' 15 challenge datasets for the evaluation. We have used the following backbone DNNs to demonstrate our approach: [24, 26, 34–36]. The details of the networks are provided on the project webpage due to a lack of space. For the microscopy images we compare against Grabcut [6], Nuclick [5], DEXTR [4] and f-BRS [3]. For the CT and MRI datasets, we have compared our method against 3-D GrabCut [33], Geos [7] and SlicSeg [8]. Table 2 shows that our technique gives an improvement in user annotation time of 13.3, 12.5, 17.8, 10.2 and 12.4 times compared to full human annotation time and 2.8, 3.0, 1.9, 4.4 and 8.6 times compared to other approaches for nucleus, multiple cells, liver and tumour, multiple organs, and brain segmentation respectively. We also compared the

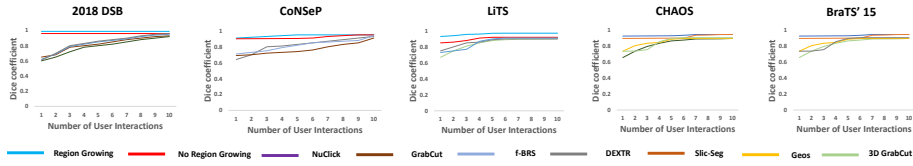


Fig. 3: Improvement in segmentation accuracy per user interaction: Our models (region and no-region growing) consistently achieve best accuracy, and in the least number of user interactions.

Table 3: **Left:** Dice Coefficient improvement for tissues with each interaction by medical expert. **Right:** User Interaction Time (**UT**) and Machine Time (**MT**) for distribution mismatch scenario (in mins).

| Tissue Type | 1 | 2 | 3 | 4 | 5 | | Method | UT | MT |
|---------------------|------|------|------|------|------|--|---------|----|----|
| Nucleus | 0.54 | 0.62 | 0.76 | 0.81 | 0.86 | | Ours | 8 | 7 |
| Healthy | 0.64 | 0.73 | 0.79 | 0.85 | 0.9 | | NuClick | 13 | 10 |
| Necrosis | 0.61 | 0.65 | 0.72 | 0.81 | 0.85 | | DEXTR | 20 | 11 |
| Edema | 0.72 | 0.75 | 0.82 | 0.89 | 0.92 | | f-BRS | 23 | 12 |
| Enhancing tumor | 0.62 | 0.65 | 0.74 | 0.85 | 0.89 | | GrabCut | 25 | 13 |
| Non-Enhancing tumor | 0.71 | 0.75 | 0.83 | 0.87 | 0.92 | | | | |
| Liver | 0.73 | 0.75 | 0.81 | 0.89 | 0.92 | | | | |
| Tumor | 0.67 | 0.72 | 0.83 | 0.87 | 0.89 | | | | |

segmentation accuracy per user interaction for every method. Fig. 3 shows that our method with region growing outperforms all the methods both in terms of accuracy achieved, and the number of iterations taken to achieve it.

Fig. 2 shows the visual results. The top row shows the segmentation obtained by adding multiple labels in one interaction by our approach. We segment both the tumors and the entire liver by using two scribbles at the same time. One of the important capabilities of our network is to add a label missing from the initial segmentation which is shown in the middle row. Note that our method does not require any pre-training with a specific backbone for interactive segmentation. This allows us to use the backbone networks that were trained for segmenting a particular organ. This ability is especially useful in the data-scarce medical setting when the DNN model for a particular organ is unavailable. This capability is demonstrated in the bottom row of Fig. 2 where a model trained for segmenting liver on LiTS challenge [27] is used to segment the heart from SegThor challenge [28].

Distribution Mismatch: The current methods cannot handle distribution mismatches forcing pre-training on each specific dataset, requiring significant time, effort, and cost. Our method does not need any pre-training. We demonstrate the advantage on the MonuSeg dataset [25] using the model pre-trained on the 2018 Data Science Bowl [24]. Table 3 (Right) shows that our method requires much less user interaction and machine time compared to other methods.

Evaluation of our method by medical experts: Our approach was tested by medical experts: two general practitioners and a radiologist. We select five most challenging images/scans from the 2018 Data Science Bowl, LiTS, and BraTS’ 15 datasets with the least dice score when segmented with the pre-trained segmentation model. The LiTS and the BraTS’ 15 datasets were selected owing to their clinical relevance for the diagnosis and volumetric estimation of tumors. Table 3 (Left) gives the dice coefficient after each interaction. The improvement in user interaction and machine time are provided in the supplementary material on the project webpage.

5 Conclusion

Modern DNNs for image segmentation require a considerable amount of annotated data for training. Our approach allows using an arbitrary DNN for segmentation and converting it to an interactive segmentation. Our experiments show that we did not require any prior training with the scribbles and yet outperform the state-of-the-art approaches, saving upto 17x (from 120 to 7 mins) in correction time for a medical resource personnel.

References

1. Project page. <http://cvit.iit.ac.in/research/projects/cvit-projects/semi-automatic-medical-image-annotation>
2. Jang, W., Kim, C.: Interactive image segmentation via backpropagating refinement scheme. In: 2019 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2019)
3. Sofiiuk, K., Petrov, I., Barinova, O., Konushin, A.: f-BRS: Rethinking backpropagating refinement for interactive segmentation. In: 2020 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2020)
4. Maninis, K., Caelles, S., Pont-Tuset, J., Van Gool, L.: Deep extreme cut: From extreme points to object segmentation. In: 2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
5. Jahanifar, M., Koohbanani, N.A., Rajpoot, N.: Nuclick: From clicks in the nuclei to nuclear boundaries. *Med Image Anal* **65** (2020)
6. Rother, C., Kolmogorov, V., Blake, A.: Grabcut interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)* **23**(3) (2004)
7. Vezhnevets, V., Konouchine, V.: Geos: Geodesic image segmentation. In: 2018 European Conference on Computer Vision (ECCV). Springer (2008)
8. Wang, G., et al.: Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. *Med Image Anal* **34** (2016)
9. Straehle, C., Köthe, U., Knott, G., Hamprecht, F.: Carving: scalable interactive segmentation of neural volume electron microscopy images. In: 2011 Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer (2011)
10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int. J. Comput. Vis.* **1** (1988)

11. Top, A., Hamarneh, G., Abugharbieh, R.: Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3d image segmentation. In: 2010 International MICCAI Workshop on Medical Computer Vision. Springer (2010)
12. Grady, L.: Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell* **28** (2006)
13. Sahoo, P.K., Soltani, S., Wong, A., Chen, Y.C.: A survey of thresholding techniques. *Comput. Gr. Image Process.* **41** (1988)
14. Horowitz, S., Pavlidis, T.: Picture segmentation by a tree traversal algorithm. *JACM* **23** (2016)
15. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: 2016 Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). IEEE (2016)
16. Li, Z., Chen, Q., Koltun, V.: Interactive image segmentation with latent diversity. In: 2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
17. Xu, N., Price, B., Cohen, S., Yang, J., Huang, T.: Deep interactive object selection. In: 2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2016)
18. Acuna, D., Ling, H., Kar, A., Fidler, S.: Efficient interactive annotation of segmentation datasets with Polygon-RNN++. In: 2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
19. Wang, G., et al.: DeepiGeos: A deep interactive geodesic framework for medical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell* **41** (2019)
20. Wang, G., et al.: Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. Med. Imaging.* **37** (2018)
21. Amrehn, M., et al.: UI-Net: Interactive artificial neural networks for iterative image segmentation based on a user model. In: Eurographics Workshop on Visual Computing for Biology and Medicine (2017)
22. Lee, H., Jeong, W.: Scribble2Label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: 2020 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer (2020)
23. Lee, J., Mehta, S., Wick, M., Tristan, J., Carbonell, J.: Gradient-based inference for networks with output constraints. In: 2019 Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). vol. 33 (2019)
24. Caicedo, J., et al.: Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16** (2019)
25. Kumar, N., et al.: A multi-organ nucleus segmentation challenge. *IEEE Trans. Med. Imaging.* **39** (2020)
26. Graham, S., et al.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal.* (2019)
27. Christ, P.: LiTS - Liver tumor segmentation challenge, (2017), <https://competitions.codalab.org/competitions/17094>
28. Petitjean, C., Ruan, S., Lambert, Z., Dubray, B.: Proceedings of the 2019 challenge on segmentation of thoracic organs at risk in ct images,. In: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA) (2019)
29. Menze, B., et al.: The multimodal brain tumor image segmentation benchmark (BRATS),. *IEEE Trans. Med. Imaging* (2014)

30. Kavur, A.E., Selver, M.A., Dicle, O., Bar, M., Gezer, N.S.: CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data. *Med Image Anal.* (2019)
31. Torralba, A., Russell, B., Yuen, J.: LabelMe: Online image annotation and applications. *Int. J. Comput. Vis.* **98**(8) (2010)
32. Pieper, S., Halle, M., Kikinis, R.: 3D Slicer. In: 2004 IEEE International Symposium on Biomedical Imaging: nano to macro (ISBI). IEEE (2004)
33. Meyer, G.P., Do, M.N.: 3D GrabCut: Interactive foreground extraction for reconstructed 3d scenes,. In: 2015 Proceedings of the Eurographics Workshop on 3D Object Retrieval. The Eurographics Association (2015)
34. Bellver, M., et al.: Detection-aided liver lesion segmentation using deep learning. *arXiv preprint arXiv:1711.11069* (2017)
35. Sinha, A., Dolz, J.: Multi-scale self-guided attention for medical image segmentation. *IEEE J Biomed Health Inform* (2020)
36. Qin, Y., et al.: Autofocus layer for semantic segmentation. In: 2018 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer (2018)

Supplementary Material

Efficient and Generic Interactive Segmentation Framework

A Ablation Study

We conducted studies to determine the effect of various hyper-parameters for our method.

1. **Optimum iterations:** We obtained the optimum number of iterations of back-propagation for obtaining a dice coefficient of 0.95 for each segmentation network. As seen in the Fig. 4, the optimum number of iterations was 80, 100, 130, 140 and 130 for the 2018 Data Science Bowl [24], CoNSeP [26], LiTS Challenge [27], CHAOS dataset [30] and BraTS' 15 [29] segmentation respectively.

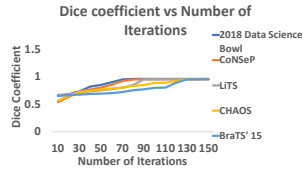


Fig. 4: Optimum Iterations Determination

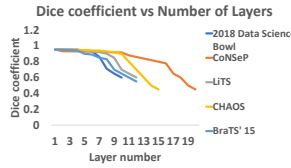


Fig. 5: Optimum Layer Determination

Table 4: Number of Mouse Clicks(N), User Time (UT) and Machine Time (MT) for various user inputs (in mins).

| Method | N | UT | MT |
|-----------------|-----|----|----|
| Point | 100 | 13 | 16 |
| Box | 34 | 10 | 15 |
| Scribble | 10 | 5 | 10 |

2. **Optimum layer:** Once, the optimum number of iterations are determined, our next step is to determine the optimum layer number upto which back-propagation needs to be performed for each segmentation network. We observe that we obtain the best possible dice coefficient for 4, 6, 4, 3 and 5 layers for the 2018 Data Science Bowl [24], CoNSeP [26], LiTS Challenge [27], CHAOS dataset [30] and BraTS' 15 [29] segmentation respectively as seen in the right panel of Fig. 5.
3. **Optimum user input type:** Our method has the unique and remarkable capability of being able to work with any type of user input such as points, boxes and scribbles. We first performed experiments to determine the most suitable user input modality for segmentation correction. We found that scribbles required the least number of user interactions (30% lesser mouse clicks), as well as user and machine time (Table 4). Hence, the experiments in the paper were done with scribbles only.
4. **Optimum scribble length:** We also evaluated the effect of scribble length while using our method. We observed that without region growing, we needed more user interactions to correct the segmentation as the scribble length

reduced. However, with region growing, there was hardly any change in the number of user interactions required as seen in Fig. 6 (obtained for LiTS challenge, similar behavior observed for other datasets, but were not able to provide owing to space restrictions).

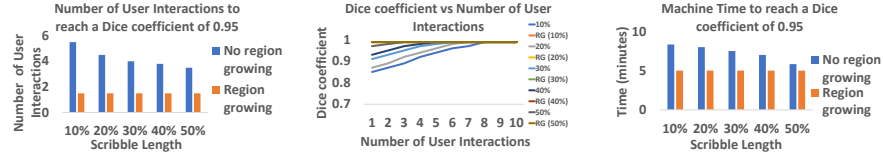


Fig. 6: Effect of scribble length on the increase in the number of user interactions (RG - Region Growing).

B Evaluation of our method by medical expert

We evaluated our interactive segmentation method with the help of medical experts. We have provided the user interaction time and machine time required for the 2018 Data Science Bowl (2018 DSB), LiTS and BraTS' 15 challenges here. It was possible to obtain a reduction in user annotation time as well as machine time as seen in Table 5.

Table 5: User Interaction Time (**UT**) and Machine Time (**MT**) in minutes for separating structures by a medical expert (**FH**: Full Human Annotation, **RG**: Our method - Region Growing, **NRG**: Our Method - No Region Growing. All the semi-automated methods [3, 4, 6-8, 33] were applied till a dice coefficient of 0.95 was reached.

| Dataset | User Interaction Time | | | | | | | | | Machine Time | | | | | | | |
|-----------|-----------------------|----|-----|-----|-----|-----|------|-----|-----|--------------|-----|-----|-----|-----|------|-----|-----|
| | FH | RG | NRG | [6] | [3] | [4] | [33] | [7] | [8] | RG | NRG | [6] | [3] | [4] | [33] | [7] | [8] |
| 2018 DSB | 55 | 4 | 6 | 15 | 14 | 14 | - | - | - | 7 | 12 | 19 | 18 | 15 | - | - | - |
| LiTS | 100 | 6 | 7 | - | - | - | 14 | 15 | 16 | 9 | 10 | - | - | - | 12 | 14 | 15 |
| BraTS' 15 | 150 | 9 | 12 | - | - | - | 65 | 75 | 90 | 50 | 80 | - | - | - | 120 | 126 | 145 |