Supplementary Materials for Regional Interactive Segmentation Networks

Jun Hao Liew[1] Yunchao Wei[2] Wei Xiong[3] Sim-Heng Ong[1,2] Jiashi Feng[2]

[1] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore

[2] Department of Electrical and Computer Engineering, National University of Singapore [3] Institute for Infocomm Research

liewjunhao@u.nus.edu {eleweiyv, eleongsh, elefjia}@nus.edu.sg wxiong@i2r.a-star.edu.sg

1. Detailed Performance Analysis:

Denote x_i and y_i as the prediction map and ground truth at location i, respectively. We propose to use the following metrics to evaluate the performance of our proposed RIS-Net without the graph cut component:

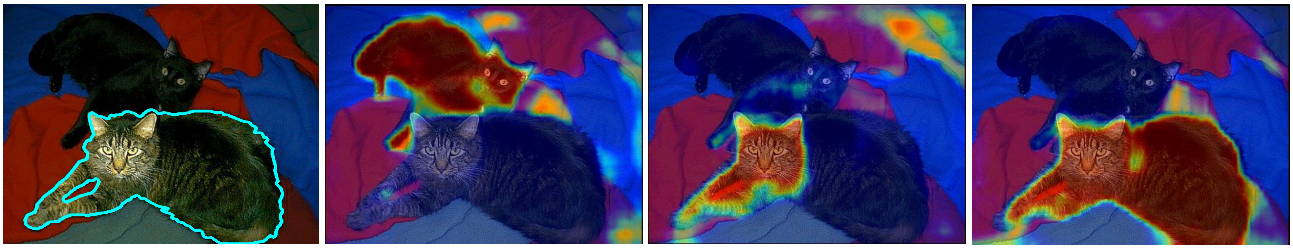
Hamming distance (HD): The Hamming distance between the ground truth and prediction map is defined as

$$HD = \frac{1}{N}\sum_{i=1}^{N} x_i \cdot (1 - y_i) + (1 - x_i) \cdot y_i$$

Coverage score (C): The coverage score is defined as

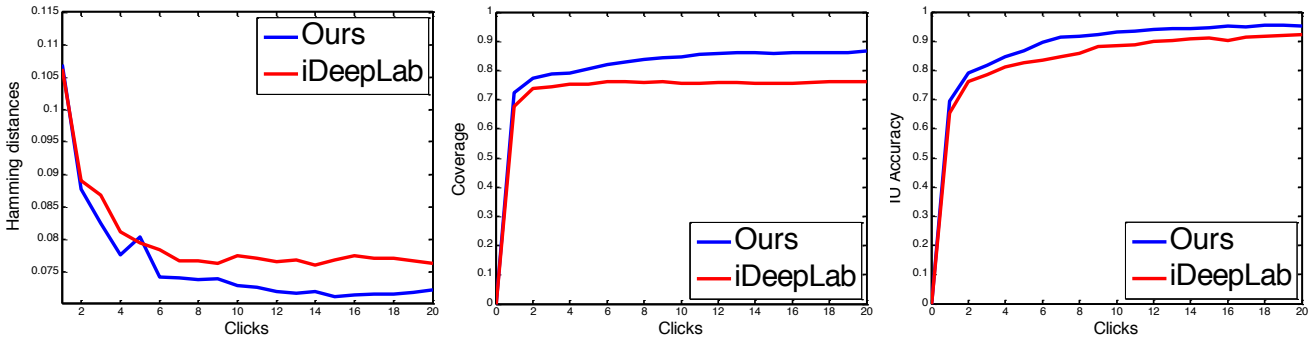
$$C = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{\sum_{i=1}^{N} y_i}$$

The Hamming distance is commonly used in the segmentation literature to measure the pixel-wise difference between two binary maps where 0 implies the two completely overlap. Here, we employ the Hamming distance metric to measure both the pixel-level accuracy and confidence of each pixel (close to 0 or 1). The smaller the Hamming distance, the more accurate is the model. In addition, we also compute the coverage score to measure the recall rate of each model where the model with higher coverage score is preferred. Figure 1 shows an intuitive example for interpreting these two metrics. The results on Grabcut and Berkeley dataset are shown in Figure 2.

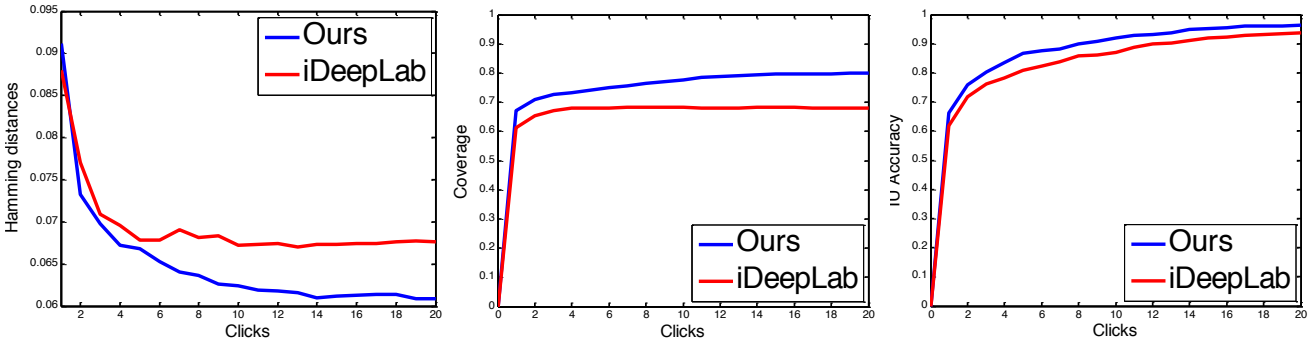


(a) Ground truth (b) HD: 0.47; C: 0.022 (c) HD: 0.27; C: 0.27 (d) HD: 0.083; C: 0.88

Figure 1: Different predictions with their corresponding Hamming distances (HD) and coverage scores (C). Three different scenarios are presented here: (b) large Hamming distance with low coverage, (c) relatively small Hamming distance but low coverage and (d) small Hamming distance with high coverage. Hamming distance is used to measure pixel-wise difference between the prediction map and the ground truth mask and it takes the confidence score on both foreground and background into consideration. A higher confidence score on target object would give smaller Hamming distance, and a lower confidence score on background would also give smaller Hamming distance. On the other hand, the coverage score is used to measure the proportion of ground truth mask covered by the prediction. It only considers the predicted confidence scores on foreground. In short, a model with low Hamming distance and high coverage score is preferred.



(a) Hamming distance (Grabcut) (b) Coverage (Grabcut) (c) IU accuracy (Grabcut)



(d) Hamming distance (Berkeley) (e) Coverage (Berkeley) (f) IU Accuracy (Berkeley)

Figure 2: The Hamming distance (the smaller the better), coverage score (the higher the better) and IU accuracy for the Grabcut and Berkeley datasets.

1.1. Discussion

Figure 2 gives the evaluation results based on Hamming distance and coverage score. As shown in Figure 2, we can see that our RIS-Net consistently outperforms the state-of-the-art iDeepLab model on both datasets with significantly smaller Hamming distance and higher coverage score. On the Grabcut dataset, we can see that the IU accuracy of iDeepLab model increases with increasing number of clicks but the coverage score is not improved. Note that the IU accuracies are evaluated on the *binary* maps that are generated by applying graph cut to the network prediction. Different from IU accuracy, the coverage score is evaluated on the prediction directly. The increasing IU accuracy with nearly constant coverage score of iDeepLab suggests that it mainly relies on graph cut optimization (for producing the binary segmentation mask) to refine its prediction. On the other hand, we can see that increasing click numbers leads to improvement of both coverage score and IU accuracy for our proposed RIS-Net benefiting from the fact that the RIS-Net is designed to fully exploit the additional user-provided information during refinement, thus accelerating the refinement progress. Similar observation can be made for the Berkeley dataset.

2. More Visual Comparisons with the State-of-the-Arts

Here, we also present more visual comparisons of our RIS-Net with the state-of-the-art solutions.

Graph cut [2]	Geodesic matting [1]	Random walk [3]	Euclidean star convexity [4]	Geodesic star convexity [4]	iDeepLab	Ours	Ground truth

Figure 3: Qualitative comparison between the baseline and our model given the same set of user interactions. The positive and negative clicks are denoted by blue dots (·) and red crosses (×) respectively. Object boundaries are highlighted in cyan. Best viewed in color.

References

[1] X. Bai and G. Sapiro. Geodesic matting: A framework for fast interactive image and video segmentation and matting. *International Journal on Computer Vision*, 82(2):113–132, 2009. 2, 3

[2] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *IEEE International Conference on Computer Vision*, volume 1, pages 105–112, 2001. 2, 3

[3] L. Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 28(11):1768–1783, 2006. 2, 3

[4] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3129–3136, 2010. 2, 3