Supplementary Material - PatchRefineNet: Improving Binary Segmentation by Incorporating Signals from Optimal Patch-wise Binarization

Savinay Nagendra Daniel Kifer
sxn265@psu.edu dkifer@cse.psu.edu
Department of Computer Science
The Pennsylvania State University
University Park

1. Choosing patch size parameter P

The patch size used by the local branch of PRN is controlled by the parameter P. Fig. 1 shows the effect of P on mIoU, which is evaluated on 100 randomly augmented images from the DeepGlobe training set. The best choice, $P = 64$, results in the local branch dividing the input logit map into sixty-four patches of size 64×64. Since the testing set was not used at all for choosing patch size, it is appropriate to use $P = 64$ in the rest of our experiments.

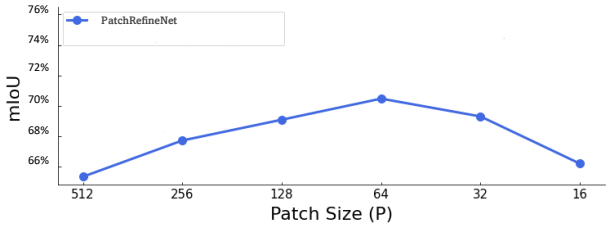

Figure 1. Effect of patch size parameter **P**. mIoU is computed using the DeepGlobe training set.

2. Comparison with DenseCRF

In this section, we compare the performance of PRN with DenseCRF [1], which is a popular method for post-processing segmentation maps.

We performed grid search on 100 randomly augmented images from DeepGlobe, Kvasir-SEG, DUTS and FSS-1000 traning sets to find the respective DenseCRF hyperparameters, as suggested in the original paper. [1]. Tables 1, 2, 3 and 4 tabulate quantitative results of performance comparison. Figure 2 shows the qualitative comparison between PRN and DenseCRF refinement. It can be observed that DenseCRF is unable to remove large regions of false positives and false negatives from the base network's segmentation maps. This is because, DenseCRF is an unsupervised method that does not capture global semantics between pixels regions. It is only able to achieve smoothing along the boundary, which is the cause for a small improvement in mIoU and mBA. We also observed that applying DenseCRF to PRN does not provide any significant improvement in performance. The results are consistent across all datasets.

DUTS [12]		
Methods	mIoU (%)	mBA (%)
RFCN [13]	52.8	40.7
(+) DenseCRF	$54.6_{\uparrow 1.8}$	$42.2_{\uparrow 1.5}$
(+) PRN	$\mathbf{57.1_{\uparrow 4.3}}$	$\mathbf{48.5_{\uparrow 7.8}}$
PFAN [15]	66.1	51.2
(+) DenseCRF	$67.0_{\uparrow 0.9}$	$52.4_{\uparrow 1.2}$
(+) PRN	$\mathbf{69.9_{\uparrow 3.8}}$	$\mathbf{58.6_{\uparrow 7.4}}$

Table 1. Quantitative results comparing PRN (P=64) with DenseCRF on DUTS saliency detection dataset.

Kvasir-SEG [5]		
Methods	mIoU (%)	mBA (%)
U-Net [10]	41.5	38.8
(+) DenseCRF	$43.2_{\uparrow 1.7}$	$41.6_{\uparrow 2.8}$
(+) PRN	$\mathbf{47.8_{\uparrow 6.3}}$	$\mathbf{46.3_{\uparrow 7.5}}$
ResUnet [3]	46.8	45.7
(+) DenseCRF	$48.3_{\uparrow 1.5}$	$47.1_{\uparrow 1.4}$
(+) PRN	$\mathbf{52.9_{\uparrow 6.1}}$	$\mathbf{52.5_{\uparrow 6.8}}$
ResUnet++ [6]	55.9	56.8
(+) DenseCRF	$56.8_{\uparrow 0.9}$	$58.2_{\uparrow 1.4}$
(+) PRN	$\mathbf{61.7_{\uparrow 5.8}}$	$\mathbf{62.9_{\uparrow 6.1}}$
SSFormer-S [11]	86.8	69.7
(+) DenseCRF	$87.6_{\uparrow 0.8}$	$70.9_{\uparrow 1.2}$
(+) PRN	$\mathbf{89.1_{\uparrow 2.3}}$	$\mathbf{72.3_{\uparrow 2.6}}$

Table 2. Quantitative results comparing PRN (P=64) with DenseCRF on Kvasir-SEG dataset.

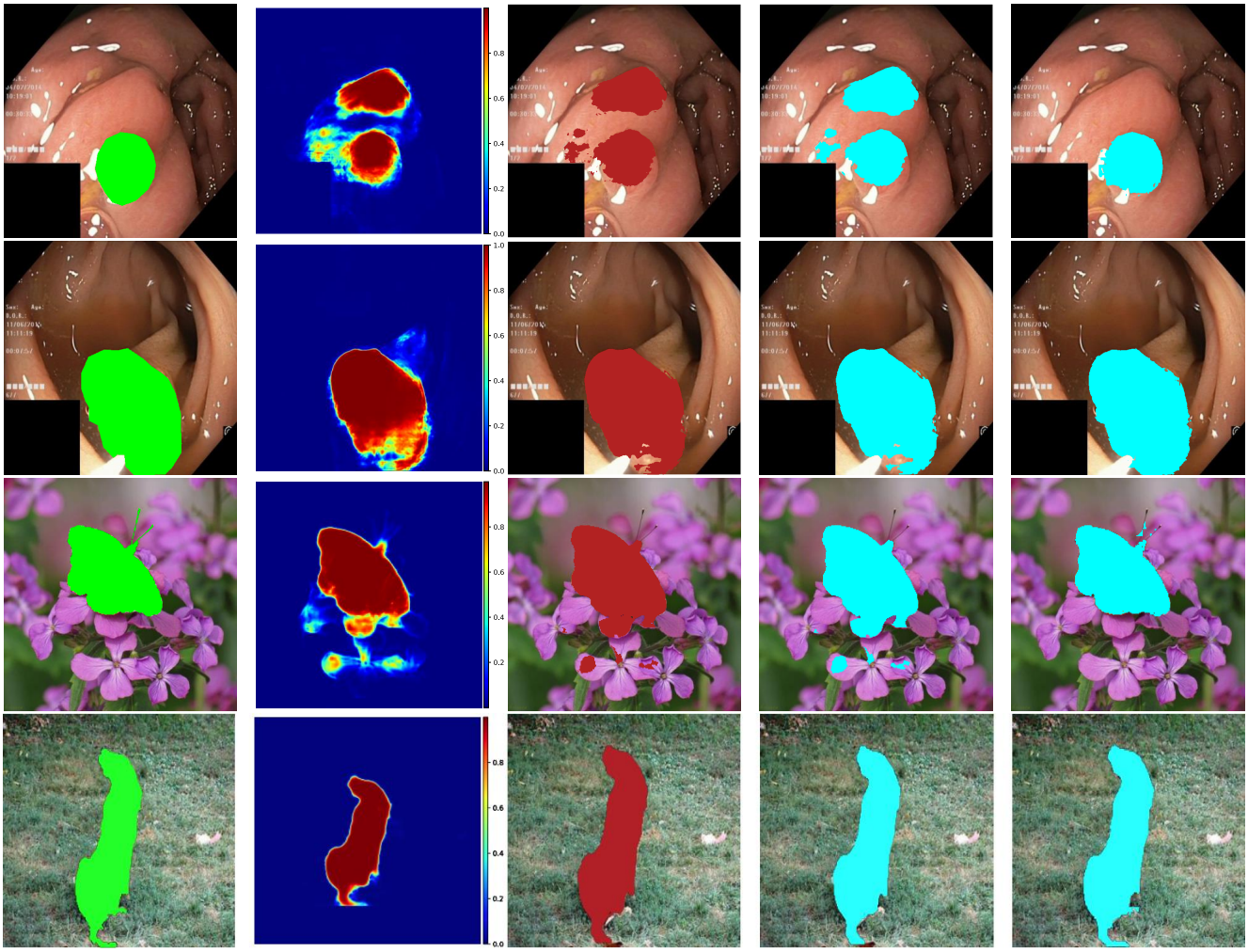

Figure 2. **Qualitative results comparing performance of PRN (P=64) with DenseCRF on samples from Kvasir-SEG (Rows 1&2) and DUTS saliency detection (Rows 3&4) test sets.** Left to Right: Ground Truth, Logit map from base network SSFormer-S [11] (Rows 1&2) and PFAN [15] (Rows 3&4), SSFormer-S prediction (Rows 1&2) and PFAN prediction (Rows 3&4), Refined by DenseCRF, Refined by PRN.

3. More qualitative results for performance evaluation of PRN

Figures 3, 4 and 5 show qualitative results for PRN refinement PRN with a patch-size parameter $P = 64$ when applied to DeepGlobe, DUTS saliency detection and Kvasir-SEG datasets respectively. Going from left to right, the figures show Ground Truth, Logit map from the base network, binarized prediction of the base network, and PRN refined map. Yellow boxes mark the regions of refinement. It can be observed that PRN is able to achieve significant visual improvement compared to the corresponding segmentation maps from state-of-the-art base networks with fixed (0.5) thresholding.

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1, 7

[2] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018. 7

[3] Foivos I Diakogiannis, François Waldner, Peter Caccetta,

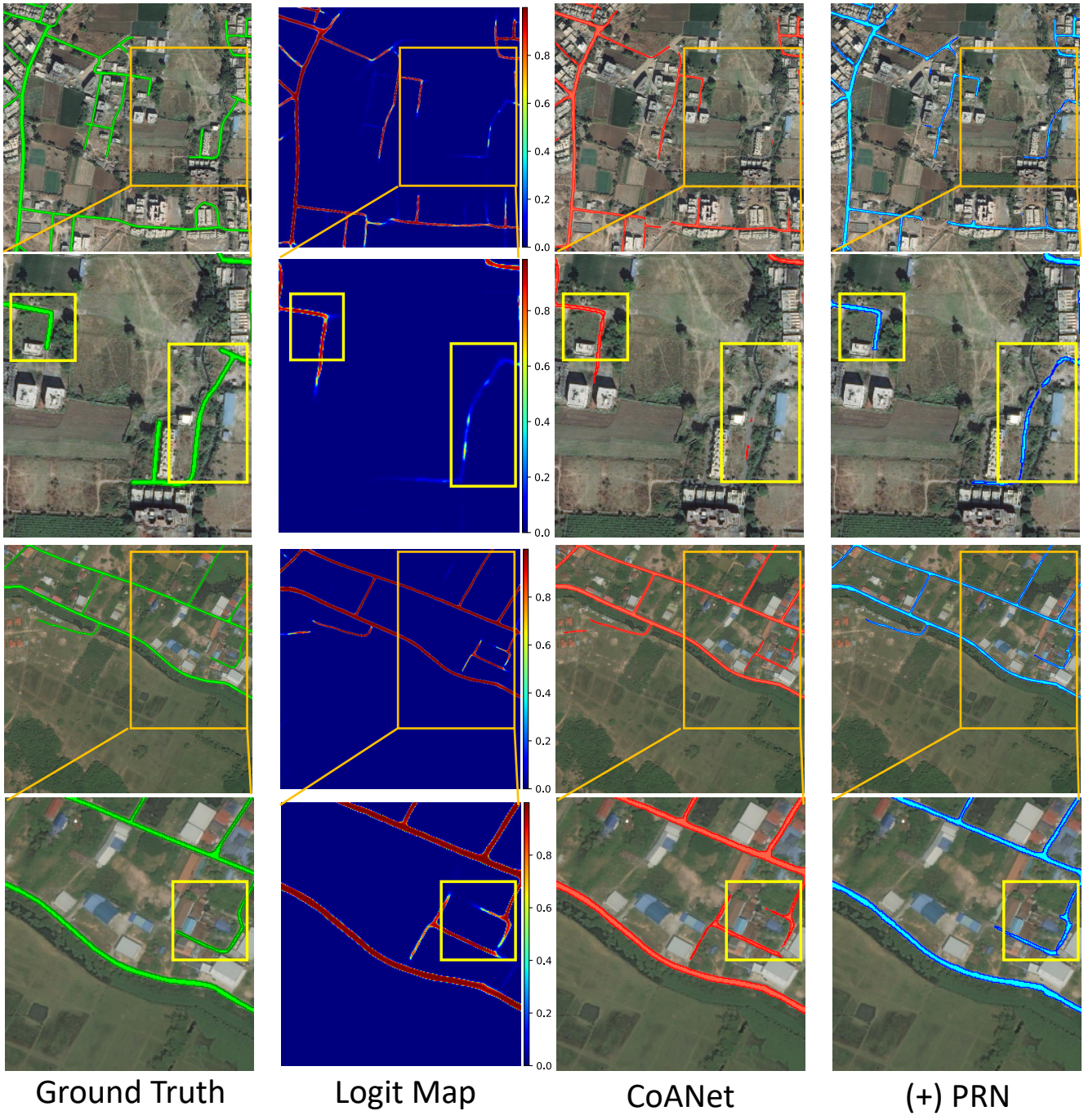

Figure 3. **Qualitative results produced by PRN** $(P = 64)$ **on samples from DeepGlobe test dataset.** Left to Right: Ground Truth, Logit map from base network CoANet [9], Prediction from CoANet , Refined by PRN. Yellow boxes denote regions of refinement.

and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020. 1

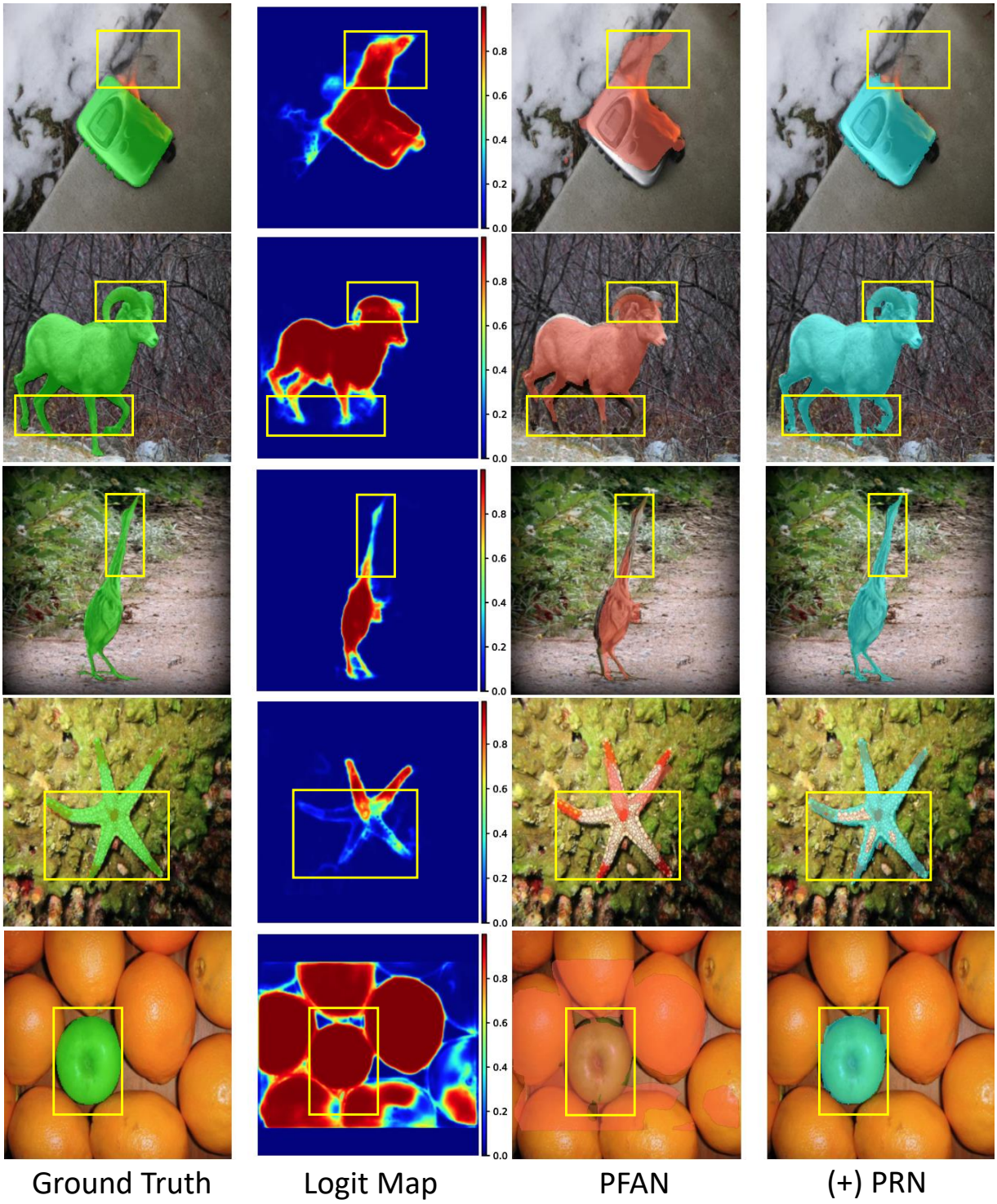

Figure 4. **Qualitative results produced by PRN** $(P = 64)$ **on samples from DUTS saliency detection test dataset.** Left to Right: Ground Truth, Logit map from base network PFAN [14], Prediction from PFAN, Refined by PRN. Yellow boxes denote regions of refinement.

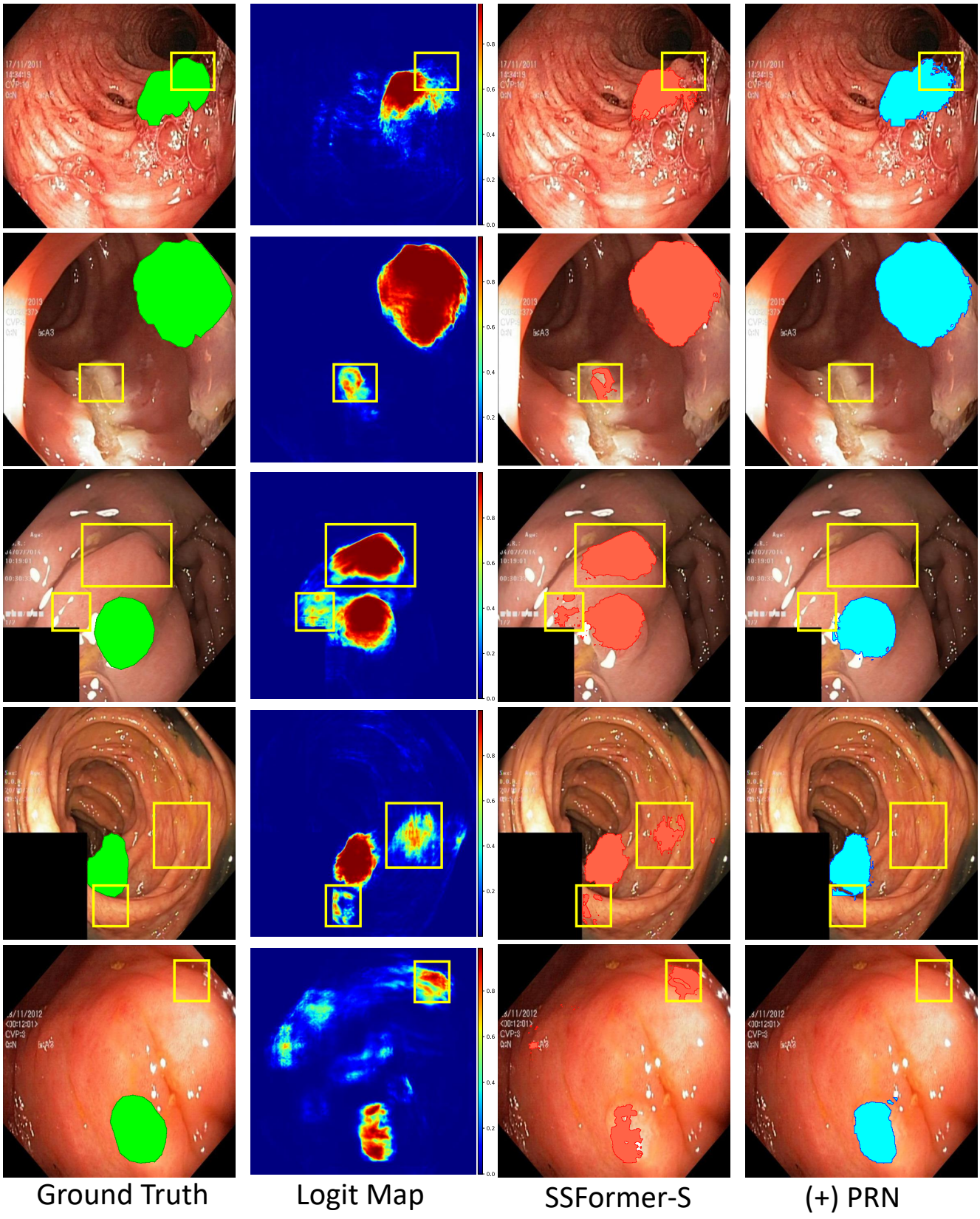

Figure 5. **Qualitative results produced by PRN** $(P = 64)$ **on samples from Kvasir-SEG test dataset.** Left to Right: Ground Truth, Logit map from base network SSFormer-S [11], Prediction from SSFormer-S, Refined by PRN. Yellow boxes denote regions of refinement.

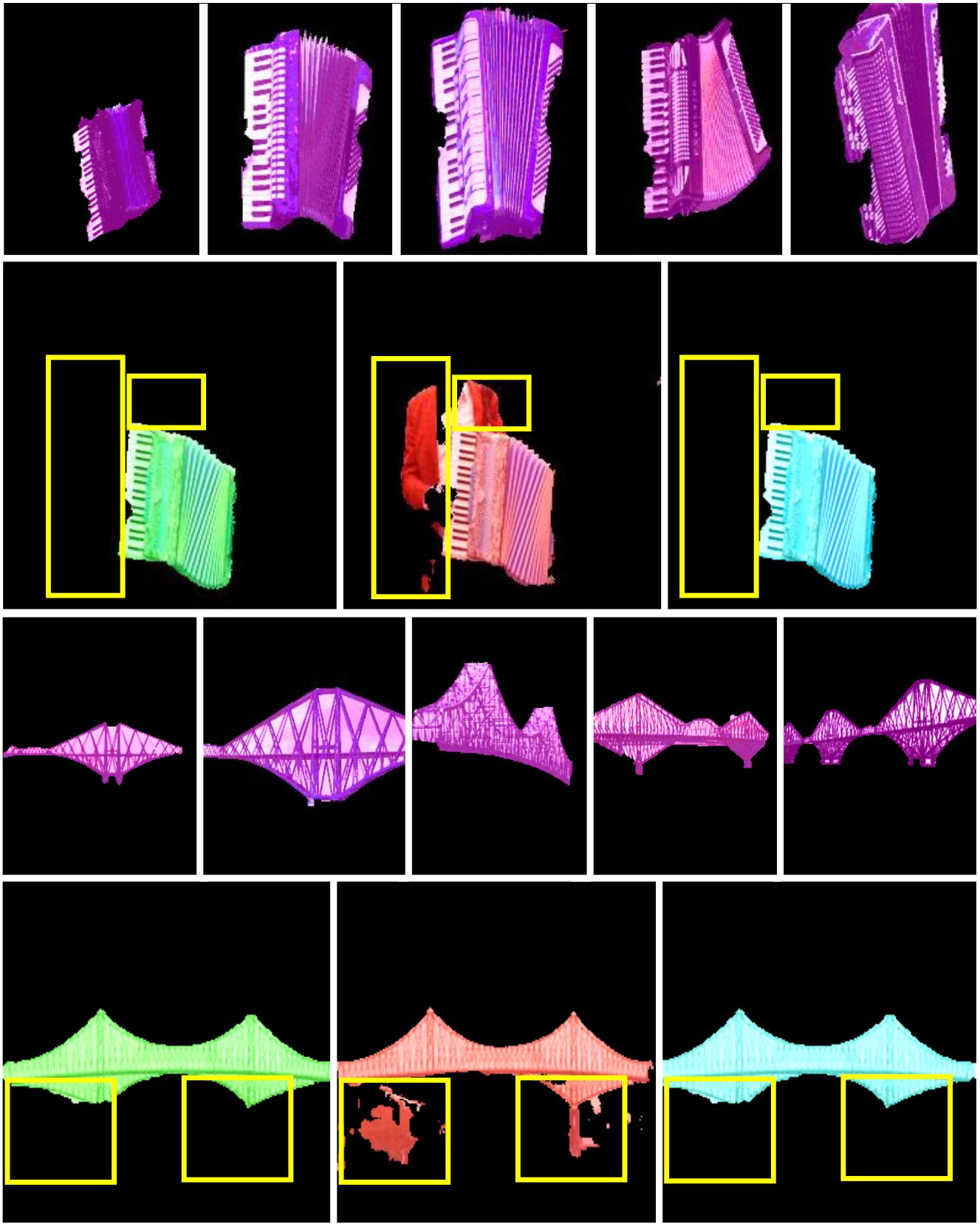

Figure 6. **Qualitative results produced by PRN** ($P = 64$) **on samples from FSS-1000 test dataset**. Rows 1& 3: Support images. Left to right: Ground Truth, Thresholded (0.5) prediction from Adapted Relation Network [15], Refined map by PRN. Yellow boxes denote regions of refinement.

DeepGlobe [2]		
Methods	mIoU (%)	mBA (%)
FCN-8s [8]	48.6	31.7
(+) DenseCRF	$50.9_{\uparrow 2.3}$	$34.9_{\uparrow 3.2}$
(+) PRN	$\mathbf{54.1_{\uparrow 5.5}}$	$\mathbf{42.3_{\uparrow 10.6}}$
U-Net [10]	55.8	37.6
(+) DenseCRF	$57.6_{\uparrow 1.8}$	$40.5_{\uparrow 2.9}$
(+) PRN	$\mathbf{60.9_{\uparrow 5.1}}$	$\mathbf{47.4_{\uparrow 9.8}}$
DeepLabV3+ [1]	59.2	47.6
(+) DenseCRF	$59.9_{\uparrow 0.7}$	$49.3_{\uparrow 1.7}$
(+) PRN	$\mathbf{61.9_{\uparrow 2.7}}$	$\mathbf{55.9_{\uparrow 8.3}}$
PSPNet [14]	59.8	48.2
(+) DenseCRF	$60.3_{\uparrow 0.5}$	$49.7_{\uparrow 1.5}$
(+) PRN	$\mathbf{62.4_{\uparrow 2.6}}$	$\mathbf{56.6_{\uparrow 8.4}}$
D-LinkNet [16]	61.3	49.8
(+) DenseCRF	$62.0_{\uparrow 0.7}$	$51.4_{\uparrow 1.6}$
(+) PRN	$\mathbf{64.4_{\uparrow 3.1}}$	$\mathbf{56.6_{\uparrow 6.8}}$
CoANet [9]	67.9	58.4
(+) DenseCRF	$69.0_{\uparrow 1.1}$	$59.6_{\uparrow 1.2}$
(+) PRN	$\mathbf{70.6_{\uparrow 2.7}}$	$\mathbf{62.1_{\uparrow 3.7}}$

Table 3. Quantitative results comparing PRN (P=64) with DenseCRF on DeepGlobe dataset.

FSS-1000 [7]		
Methods	mIoU (%)	mBA (%)
Adapted Relation Network [7]	80.1	69.8
(+) DenseCRF	$81.2_{\uparrow 1.1}$	$71.1_{\uparrow 1.3}$
(+) PRN	$\mathbf{82.7_{\uparrow 2.6}}$	$\mathbf{72.9_{\uparrow 3.1}}$
EfficientLab [4]	82.8	71.1
(+) DenseCRF	$83.8_{\uparrow 1.0}$	$72.3_{\uparrow 1.2}$
(+) PRN	$\mathbf{84.1_{\uparrow 1.3}}$	$\mathbf{73.2_{\uparrow 2.1}}$

Table 4. Quantitative results comparing PRN (P=64) with DenseCRF on FSS-1000 dataset.

[4] Sean M Hendryx, Andrew B Leach, Paul D Hein, and Clayton T Morrison. Meta-learning initializations for image segmentation. *arXiv preprint arXiv:1912.06290*, 2019. 7

[5] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*, pages 451–462. Springer, 2020. 1

[6] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255. IEEE, 2019. 1

[7] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang. Fss-1000: A 1000-class dataset for few-shot segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2869–2878, 2020. 7

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 7

[9] Jie Mei, Rou-Jing Li, Wang Gao, and Ming-Ming Cheng. Coanet: Connectivity attention network for road extraction from satellite imagery. *IEEE Transactions on Image Processing*, 30:8540–8552, 2021. 3, 7

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 7

[11] Jinfeng Wang, Qiming Huang, Feilong Tang, Jia Meng, Jionglong Su, and Sifan Song. Stepwise feature fusion: Local guides global. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III*, pages 110–120. Springer, 2022. 1, 2, 5

[12] Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 136–145, 2017. 1

[13] Linzhao Wang, Lijun Wang, Huchuan Lu, Pingping Zhang, and Xiang Ruan. Saliency detection with recurrent fully convolutional networks. In *European conference on computer vision*, pages 825–841. Springer, 2016. 1

[14] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 4, 7

[15] Ting Zhao and Xiangqian Wu. Pyramid feature attention network for saliency detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3085–3094, 2019. 1, 2, 6

[16] Lichen Zhou, Chuang Zhang, and Ming Wu. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 182–186, 2018. 7