

# Supplementary Material: Cross-patch Dense Contrastive Learning for Semi-supervised Segmentation of Cellular Nuclei in Histopathologic Images

Huisi Wu<sup>1</sup>, Zhaoze Wang<sup>1</sup>, Youyi Song<sup>2</sup>, Lin Yang<sup>2</sup>, Jing Qin<sup>2</sup>

<sup>1</sup>Shenzhen University, <sup>2</sup>The Hong Kong Polytechnic University

## 1. More Hyper Parameter Analysis

**Different Weights for Contrastive Loss.** The influence of our proposed cross-patch dense contrastive learning module depends on the weight ( $w_{contr}$ ) of contrastive loss ( $L_{contr}$ ). With low  $w_{contr}$ , the network cannot efficiently exert the effect of contrastive learning, while high  $w_{contr}$  may corrupt the optimization of learning rate. Therefore, the weight setting for contrastive loss is very important. Table 1 shows our ablation study on different  $w_{contr}$ . We can observe that, while the  $w_{contr}$  is set as 0.1, our method achieves the best performance on two histopathologic datasets.

Dataset	$w_{contr}$	DC(%)	JC(%)	ACC(%)	SP(%)	SE(%)
DSB	0	86.40	77.68	96.84	98.06	87.06
	0.01	86.61	78.25	96.84	97.85	88.22
	0.05	86.98	78.74	96.99	98.12	<b>89.20</b>
	<b>0.1</b>	<b>87.49</b>	<b>79.35</b>	<b>96.99</b>	<b>98.38</b>	88.43
	0.5	86.12	77.18	96.53	98.33	86.13
	1	85.30	76.40	96.16	97.80	87.59
MoNuSeg	0	75.16	60.73	89.47	90.82	84.29
	0.01	75.28	60.98	89.20	89.62	<b>87.85</b>
	0.05	75.71	61.46	89.74	90.96	85.15
	<b>0.1</b>	<b>75.97</b>	<b>61.77</b>	<b>89.83</b>	<b>91.03</b>	85.36
	0.5	73.98	59.30	88.45	89.20	85.82
	1	73.96	59.31	88.82	90.32	83.07

Table 1. Statistical comparison of our ablation study for different  $w_{contr}$  applied on DSB [1] and MoNuSeg [5] datasets with 1/32 labeled training data.

**Different Patch Sizes for Contrastive Learning.** Patch-wise sampling is the primary task of our cross-patch dense contrastive learning module, where we take the advantage of inter-patch disparity as the basis for sampling negative pairs. The size of patch may change the network’s judgement of inter-patch disparity, and further affect the performance of contrastive learning. The ablation study results of different patch sizes are as shown in Table 2, which indicates that the most suitable patch size for our module

performing patch-wise sampling and feature alignment is  $\frac{h}{8} \times \frac{w}{8}$  pixels in the image space.

Dataset	size	DC(%)	JC(%)	ACC(%)	SP(%)	SE(%)
DSB	$\frac{h}{2} \times \frac{w}{2}$	86.72	78.34	96.60	97.82	88.59
	$\frac{h}{4} \times \frac{w}{4}$	86.90	78.55	96.51	97.66	89.63
	$\frac{h}{8} \times \frac{w}{8}$	<b>87.49</b>	<b>79.35</b>	<b>96.99</b>	<b>98.38</b>	88.43
	$\frac{h}{16} \times \frac{w}{16}$	87.08	78.86	96.58	97.70	<b>89.89</b>
MoNuSeg	$\frac{h}{2} \times \frac{w}{2}$	74.26	59.46	89.65	<b>92.54</b>	78.11
	$\frac{h}{4} \times \frac{w}{4}$	75.08	60.64	89.37	90.60	84.69
	$\frac{h}{8} \times \frac{w}{8}$	<b>75.97</b>	<b>61.77</b>	<b>89.83</b>	91.03	85.36
	$\frac{h}{16} \times \frac{w}{16}$	75.14	60.75	89.32	90.32	<b>85.67</b>

Table 2. Statistical comparison of our ablation study for different patch sizes applied on DSB and MoNuSeg datasets with 1/32 labeled training data.  $h$  and  $w$  represent the height and width of the input image.

## 2. Comparison of Different Baseline Networks

We also adopted U-Net [10], with backbone ResNet-50 [3], as the baseline segmentation network and conducted experiments with different amounts of training labeled data. Statistical results in Table 3 demonstrate that our method is still effective, without over relying on a specific baseline network.

## 3. More Visual Comparisons

We provide more visual comparison with FullSup, SupOnly and other three state-of-the-art methods including TCMSv2 [8], CutMix [2], GCT [4], CCT [9] and CAC [6], on DSB and MoNuSeg datasets, with DenseUNet [7] as the base segmentation network, as shown in Figure 1.

## References

- [1] Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, et al. Nucleus segmentation across imaging experiments: the 2018

DSB							MoNuSeg						
Labeled	Method	DC(%)	JC(%)	ACC(%)	SP(%)	SE(%)	Labeled	Method	DC(%)	JC(%)	ACC(%)	SP(%)	SE(%)
1/32	SupOnly	82.22	72.29	95.61	97.13	83.19	1/32	SupOnly	70.85	55.20	88.05	<b>91.27</b>	74.90
	<b>Ours</b>	<b>85.79</b>	<b>76.68</b>	<b>96.39</b>	<b>97.84</b>	<b>87.01</b>		<b>Ours</b>	<b>74.79</b>	<b>60.33</b>	<b>89.52</b>	91.21	<b>82.93</b>
1/16	SupOnly	84.95	75.58	96.69	97.96	85.94	1/16	SupOnly	71.90	56.83	88.59	91.34	78.04
	<b>Ours</b>	<b>87.02</b>	<b>78.41</b>	<b>97.08</b>	<b>98.28</b>	<b>86.08</b>		<b>Ours</b>	<b>75.19</b>	<b>60.85</b>	<b>89.88</b>	<b>92.51</b>	<b>79.80</b>
1/8	SupOnly	85.82	76.86	96.96	<b>98.64</b>	83.36	1/8	SupOnly	74.53	60.23	90.03	<b>93.36</b>	77.87
	<b>Ours</b>	<b>87.78</b>	<b>79.49</b>	<b>97.28</b>	98.33	<b>87.89</b>		<b>Ours</b>	<b>76.49</b>	<b>62.39</b>	<b>90.38</b>	92.48	<b>82.19</b>
100%	FullSup	88.51	80.07	97.40	98.89	86.82	100%	FullSup	77.59	63.78	91.41	94.78	78.14

Table 3. Statistical results of our method on DSB and MoNuSeg dataset by equipping U-Net as the baseline segmentation network.

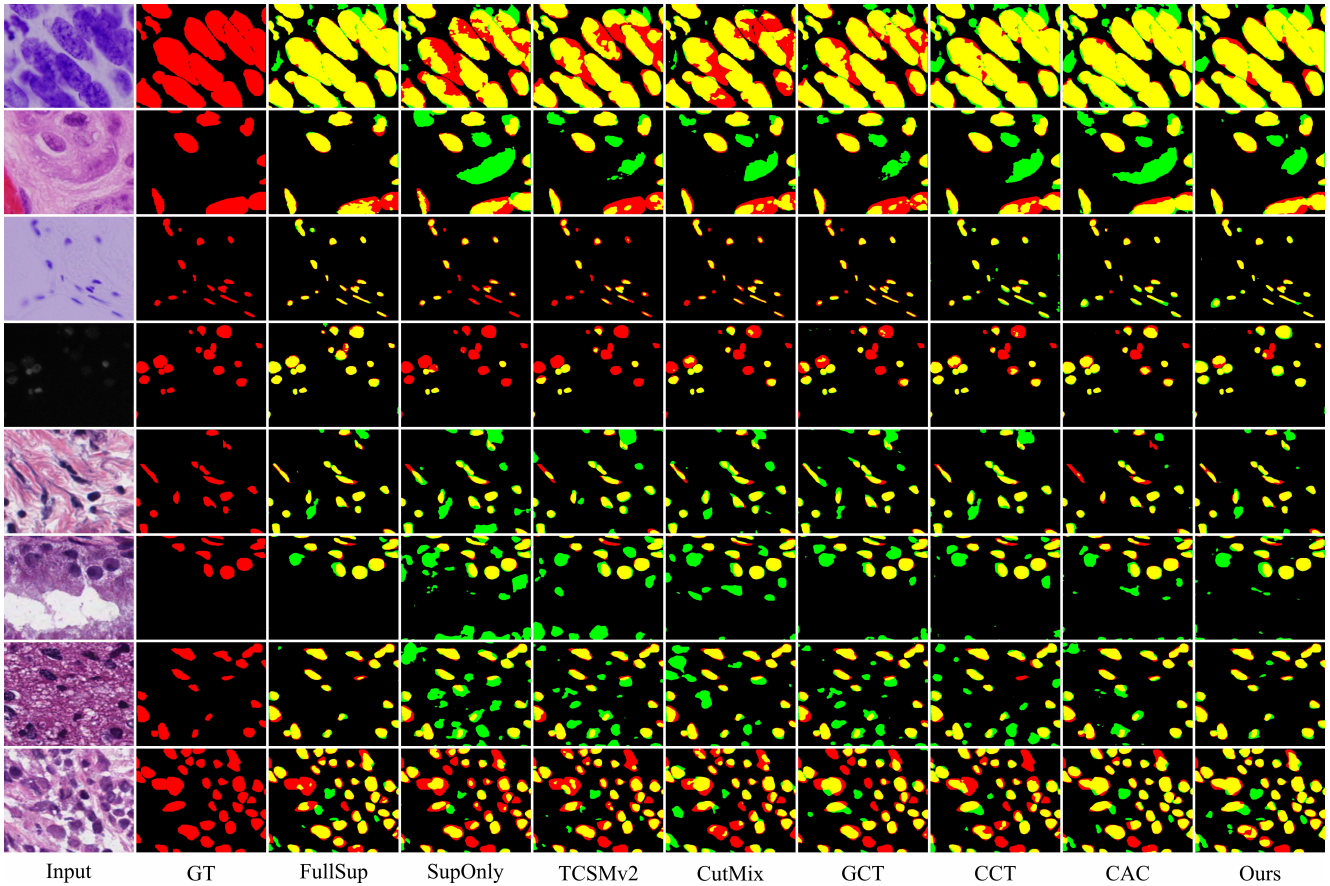


Figure 1. More visual comparison with different state-of-the-art methods in nuclei image segmentation. FullSup is trained with 100% labeled data while SupOnly with only 1/32. Other methods are trained in a semi-supervised manner with 1/32 labeled and 31/32 unlabeled data. Green and red pixels indicate the predictions and ground truth respectively while yellow pixels represent their overlap regions.

- data science bowl. *Nature methods*, 16(12):1247–1253, 2019. **1**
- [2] Geoff French, Samuli Laine, Timo Aila, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, varied perturbations. *arXiv preprint arXiv:1906.01916*, 2019. **1**
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. **1**
- [4] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise

- semi-supervised learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 429–445. Springer, 2020. [1](#)
- [5] Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng-Ann Heng, Jiahui Li, Zhiqiang Hu, et al. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 39(5):1380–1391, 2019. [1](#)
- [6] Xin Lai, Zhuotao Tian, Li Jiang, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1205–1214, 2021. [1](#)
- [7] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018. [1](#)
- [8] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, Lei Xing, and Pheng-Ann Heng. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):523–534, 2020. [1](#)
- [9] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. [1](#)
- [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](#)