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

Huisi Wu 1, Zhaoze Wang 1, Youyi Song 2, Lin Yang 2, Jing Qin 2
1 Shenzhen University, 2 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 \( w_{\text{contr}} \) of contrastive loss \( L_{\text{contr}} \). With low \( w_{\text{contr}} \), the network cannot efficiently exert the effect of contrastive learning, while high \( w_{\text{contr}} \) 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 \( w_{\text{contr}} \). We can observe that, while the \( w_{\text{contr}} \) is set as 0.1, our method achieves the best performance on two histopathologic datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>( w_{\text{contr}} )</th>
<th>DC(%)</th>
<th>JC(%)</th>
<th>ACC(%)</th>
<th>SP(%)</th>
<th>SE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSB</td>
<td>0</td>
<td>86.40</td>
<td>77.68</td>
<td>96.84</td>
<td>98.06</td>
<td>87.06</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>86.61</td>
<td>78.25</td>
<td>96.84</td>
<td>97.85</td>
<td>88.22</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>86.98</td>
<td>78.74</td>
<td>96.99</td>
<td>98.12</td>
<td>89.20</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>87.49</td>
<td>79.35</td>
<td><strong>96.99</strong></td>
<td><strong>98.38</strong></td>
<td><strong>88.43</strong></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>87.30</td>
<td>79.60</td>
<td>97.16</td>
<td>98.55</td>
<td>88.75</td>
</tr>
<tr>
<td>MoNuSeg</td>
<td>0</td>
<td>75.16</td>
<td>60.73</td>
<td>88.47</td>
<td>90.82</td>
<td>84.29</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>75.28</td>
<td>60.98</td>
<td>89.20</td>
<td>89.62</td>
<td><strong>87.85</strong></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>75.71</td>
<td>61.46</td>
<td>89.74</td>
<td>90.96</td>
<td>85.15</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td><strong>75.97</strong></td>
<td><strong>61.77</strong></td>
<td><strong>89.83</strong></td>
<td><strong>91.03</strong></td>
<td>85.36</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>73.98</td>
<td>59.30</td>
<td>88.45</td>
<td>89.20</td>
<td>85.82</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>73.96</td>
<td>59.31</td>
<td>88.82</td>
<td>90.32</td>
<td>83.07</td>
</tr>
</tbody>
</table>

Table 1. Statistical comparison of our ablation study for different \( w_{\text{contr}} \) 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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>size</th>
<th>DC(%)</th>
<th>JC(%)</th>
<th>ACC(%)</th>
<th>SP(%)</th>
<th>SE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSB</td>
<td>( \frac{h}{2} \times \frac{w}{2} )</td>
<td>86.72</td>
<td>78.34</td>
<td>96.60</td>
<td>97.82</td>
<td>88.59</td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{4} \times \frac{w}{4} )</td>
<td>86.90</td>
<td>78.55</td>
<td>96.51</td>
<td>97.66</td>
<td>89.63</td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{8} \times \frac{w}{8} )</td>
<td><strong>87.49</strong></td>
<td><strong>79.35</strong></td>
<td><strong>96.99</strong></td>
<td><strong>98.38</strong></td>
<td><strong>88.43</strong></td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{16} \times \frac{w}{16} )</td>
<td>87.08</td>
<td>78.86</td>
<td>96.58</td>
<td>97.70</td>
<td><strong>89.89</strong></td>
</tr>
<tr>
<td>MoNuSeg</td>
<td>( \frac{h}{2} \times \frac{w}{2} )</td>
<td>74.26</td>
<td>59.46</td>
<td>89.65</td>
<td><strong>92.54</strong></td>
<td>78.11</td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{4} \times \frac{w}{4} )</td>
<td>74.29</td>
<td>60.71</td>
<td>89.37</td>
<td>90.60</td>
<td>84.69</td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{8} \times \frac{w}{8} )</td>
<td><strong>75.97</strong></td>
<td><strong>61.77</strong></td>
<td><strong>89.83</strong></td>
<td><strong>91.03</strong></td>
<td>85.36</td>
</tr>
<tr>
<td></td>
<td>( \frac{h}{16} \times \frac{w}{16} )</td>
<td>75.14</td>
<td>60.75</td>
<td>89.32</td>
<td>90.32</td>
<td><strong>85.67</strong></td>
</tr>
</tbody>
</table>

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 TCSMv2 [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

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

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

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.

[4] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise


