**Table S2.** Metrics obtained in the CVC-EndoSceneStill test set for the different models and datasets used during training. Best value per metric is indicated in bold.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset for training/validation set** | **Network** | **CVC-EndoSceneStill** | | | | | | |
| Accuracy | Precision | Recall | Specificity | F2-score | Jaccard | Dice |
| CVC-EndoSceneStill | U-Net+VGG16 | 92.77±10.50 | 77.84±34.17 | 63.02±39.00 | 99.04±2.12 | 62.98±37.91 | 56.80±36.09 | 64.11±36.83 |
| U-Net+Densenet121 | 94.10±10.53 | **88.87±22.15** | 72.46±36.84 | **99.38±1.55** | 72.59±35.67 | 67.31±34.47 | 73.63±33.60 |
| Linknet+VGG16 | 93.58±10.24 | 85.05±26.50 | 69.40±36.28 | 99.36±1.25 | 69.23±34.84 | 62.77±33.81 | 70.39±33.03 |
| Linknet+Densenet121 | 94.83±9.60 | 87.88±21.73 | 79.22±32.39 | 99.00±2.15 | 78.54±31.19 | **72.16±30.93** | 78.61±29.48 |
| Kvasir-SEG | U-Net+VGG16 | 91.76±10.36 | 55.43±47.53 | 34.56±40.81 | 99.36±2.63 | 35.15±40.66 | 32.44±38.79 | 37.11±41.12 |
| U-Net+Densenet121 | 94.74±9.94 | 83.54±25.84 | **83.46±28.83** | 97.65±7.17 | **80.10±28.71** | 71.82±29.87 | **78.78±28.14** |
| Linknet+VGG16 | 93.71±9.62 | 79.03±33.46 | 64.45±39.17 | 99.19±1.68 | 64.52±38.03 | 58.62±36.18 | 65.69±36.70 |
| Linknet+Densenet121 | **95.32±9.07** | 82.98±27.51 | 76.72±34.88 | 98.77±3.83 | 75.79±33.99 | 69.10±32.53 | 75.55±32.58 |
| PICCOLO | U-Net+VGG16 | 91.03±10.99 | 66.18±38.07 | 56.96±40.54 | 96.91±6.10 | 55.73±39.19 | 47.76±36.46 | 55.43±37.98 |
| U-Net+Densenet121 | 93.86±9.97 | 79.31±33.02 | 68.99±36.56 | 99.03±2.45 | 68.76±35.75 | 62.41±34.78 | 69.46±35.13 |
| Linknet+VGG16 | 92.78±10.45 | 78.42±32.36 | 61.14±38.17 | 98.74±3.02 | 60.99±37.00 | 54.58±35.63 | 62.46±36.01 |
| Linknet+Densenet121 | 93.77±10.24 | 84.35±25.26 | 71.88±34.26 | 98.85±2.86 | 71.05±33.40 | 64.18±33.04 | 71.66±32.98 |