Tabel 1、Sartorius Cell Segmentation Ablation Results

| No | Backbone | Training Iteration | Model Ensemble | Imagenet Normalization | Color Augmentation | Test Time Augmentation | mAP |
|---|---|---|---|---|---|---|---|
| 1 | ResNet50 | 4,000 | | | | | 0.285 |
| 2 | ResNet50 + ResNeXt101 | 4,000 | V | | | | 0.291 |
| 3 | ResNet50 + ResNeXt101 | 4,000 | V | V | | | 0.293 |
| 4 | ResNet50 + ResNeXt101 | 4,000 | V | | V | | 0.288 |
| 5 | ResNet50 + ResNeXt101 | 4,000 | V | | | V | 0.25 |

From the table, it is evident that different models may learn different information. Utilizing model ensembling to combine the predictions of two models results in a slight improvement in mAP compared to using a single model. However, during testing, the introduction of multiple augmentations to generate diverse Test Time Augmentations (TTA) from input images leads to a significant decrease in mAP. This phenomenon could be attributed to the dataset's characteristics, where cells are small and densely packed. The varied augmentations in TTA may produce overlapping detections that are excluded during fusion, causing a decline in accuracy.

In addition, normalizing input images using the pixel mean and pixel standard deviation from ImageNet versus those from the dataset itself shows no significant difference. In fact, using ImageNet's pixel mean and pixel standard deviation yields a slightly higher mAP. Unexpectedly, introducing color variations as part of the augmentation during training results in a minor decrease in mAP. This discrepancy between experimental results and initial expectations may be due to the relatively uniform and limited color variation in the cell segmentation dataset. Consequently, the choice of normalization appears to have minimal impact on the network's learning, while augmentations introducing color variations create data with distributions differing from the dataset, thereby affecting model accuracy.

Table 2、Sartorius Cell Segmentation Results

| No | Backbone | Training Iterarion | Model Ensemble | mAP |
|---|---|---|---|---|
| 1 | ResNet50 | 9,679 | | 0.300 |
| 2 | ResNet50 | 10,679 | | 0.302 |
| 3 | ResNet50 + ResNet50 | 10,679 | V | 0.303 |
| 5 | ResNet50 + ResNet50 | 11,679 | V | 0.304 |

Normalization of input images was performed using the pixel mean and standard deviation from ImageNet. Two models were separately trained over an extended period, each employing a lightweight ResNet-50 as the backbone for Mask R-CNN. During testing, a model ensemble approach was adopted, combining the detection results of the two models. It is important to note that due to constraints in the competition environment, both the storage space and computational time were limited, influencing the choice of a lightweight backbone and the overall training duration for the models.