

Method	85	87.5	90	92.5	95
OLED (Ours) s_{rec}	0.63836	0.6622	0.6432	0.6426	0.6442
OLED (Ours) s_{mask}	0.64253	0.6711	0.6550	0.6595	0.6562
OLED (Ours) s_{avg}	0.6424	0.6683	0.6423	0.6573	0.6493
OLED (Ours) s_{cont}	0.6457	0.6673	0.6330	0.6580	0.6309

Table 1. Novelty Detection Average AUCROC on CIFAR-10 image dataset for different values of hyperparameter t .

Method	L_{tot}	$L_{rec} + L_{mask}$	$L_{mask} + L_{mask}$	L_{mask}
OLED (Ours) s_{rec}	0.6622	0.6499	0.6503	0.6367
OLED (Ours) s_{mask}	0.6711	0.6560	0.6529	0.6502
OLED (Ours) s_{avg}	0.6683	0.6375	0.6605	0.6242
OLED (Ours) s_{cont}	0.6673	0.6232	0.6624	0.6192

Table 2. Novelty Detection Average AUCROC on CIFAR-10 image dataset for different objectives

1. Threshold Ablation

In order to determine the sensitivity of the proposed method to changes in the threshold hyperparameter t , we evaluate the performance of the method at different settings of t . Namely, $t \in 85, 87.5, 90, 92.5, 95$ and all other hyperparameters are held constant. Specifically, the CIFAR-10 dataset is used with the experiment setup outlined in Section 4.3. The results from this experiment are available in Table 1. The performance is consistent across different values of t . Accordingly, the proposed method is robust to small changes in the t hyperparameter.

2. Loss Ablation

In this ablation study, we explore the effect of the individual components of the objective: (L_{mask} , L_{cont} and L_{rec}). To accomplish this, we evaluate the performance of different objectives on the CIFAR-10 dataset using the same experiment setup outlined in Section 4.2. The results from this experiment are available in Table 2. Although, the performance is relatively high across different objectives, it is clear that the best performance is attained when all of the components are included in the objective. In all cases, anomaly scores perform stronger if the corresponding loss is included in the objective.