Masked Attention ConvNeXt Unet with Multi-Synthesis Dynamic Weighting for Anomaly Detection and Localization

BMVC 2023

Shih-Chih Lin, Ho-Weng Lee, Yu-Hsuan Hsieh, Cheng-Yu Ho

Advisor: Dr. Shang-Hong Lai

National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT

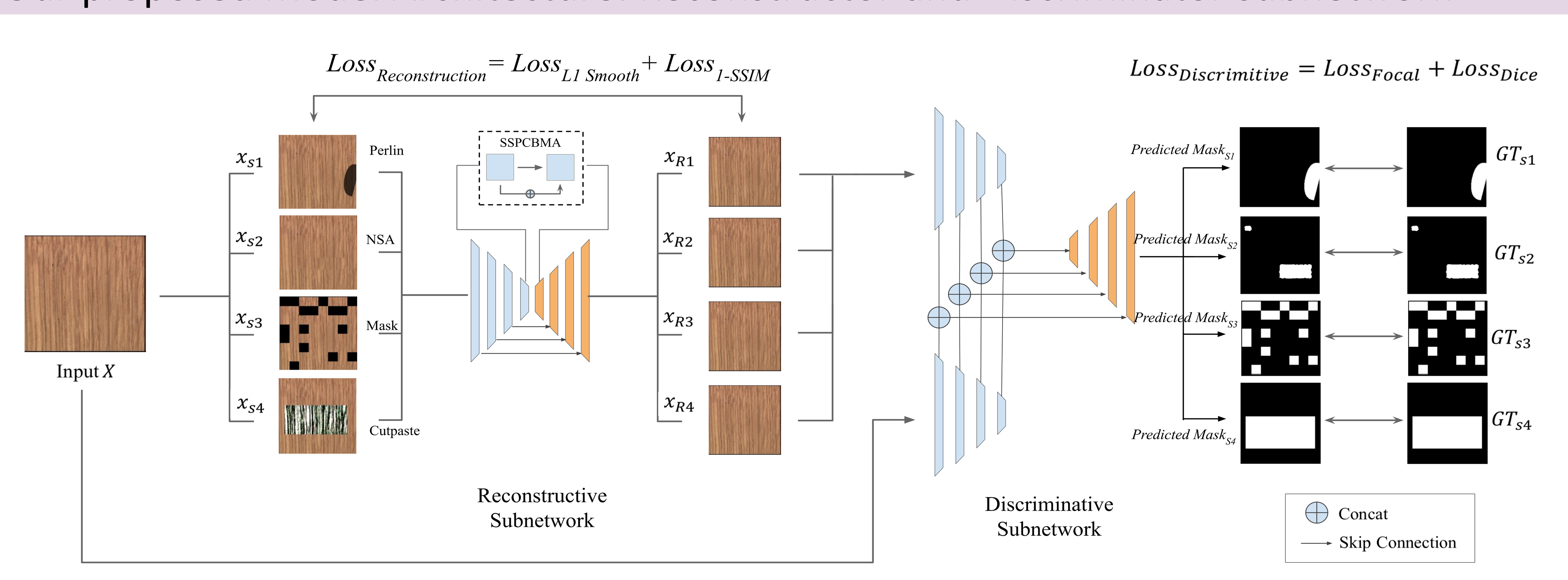
Our study introduces a novel multi-synthesis weighting strategy, denoted as MSdW, aimed at harnessing the advantage of diverse data synthesis strategies. We also construct a model architecture comprising reconstructive and discriminative subnetworks built upon the U-Net architecture with a ConvNextV2 base. We conduct a comprehensive evaluation of our proposed model across various datasets for the tasks of anomaly detection and segmentation. Notably, the datasets used for evaluation include MVTecAD, BTAD, and KSDD2. Our experimental results demonstrate that our model surpasses existing state-of-the-art methods, exhibiting significant improvements in Pixel AP and PRO indices.

REFERENCE

1. Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 8330–8339, 2021.
2. Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. arXiv preprint arXiv:2301.00808, 2023.
3. Nicolae-Cat ˘ alin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz ˘Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13576–13586, 2022
4. Rick Groenendijk, Sezer Karaoglu, Theo Gevers, and Thomas Mensink. Multi-loss weighting with coefficient of variations. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 1469–1478, 2021.

Proposed Method

Our proposed Model Architecture: Reconstructor and Discriminator subnetwork



*M*ulti-*S*ynthesis *D*ynamic *W*eighting (MSdW)

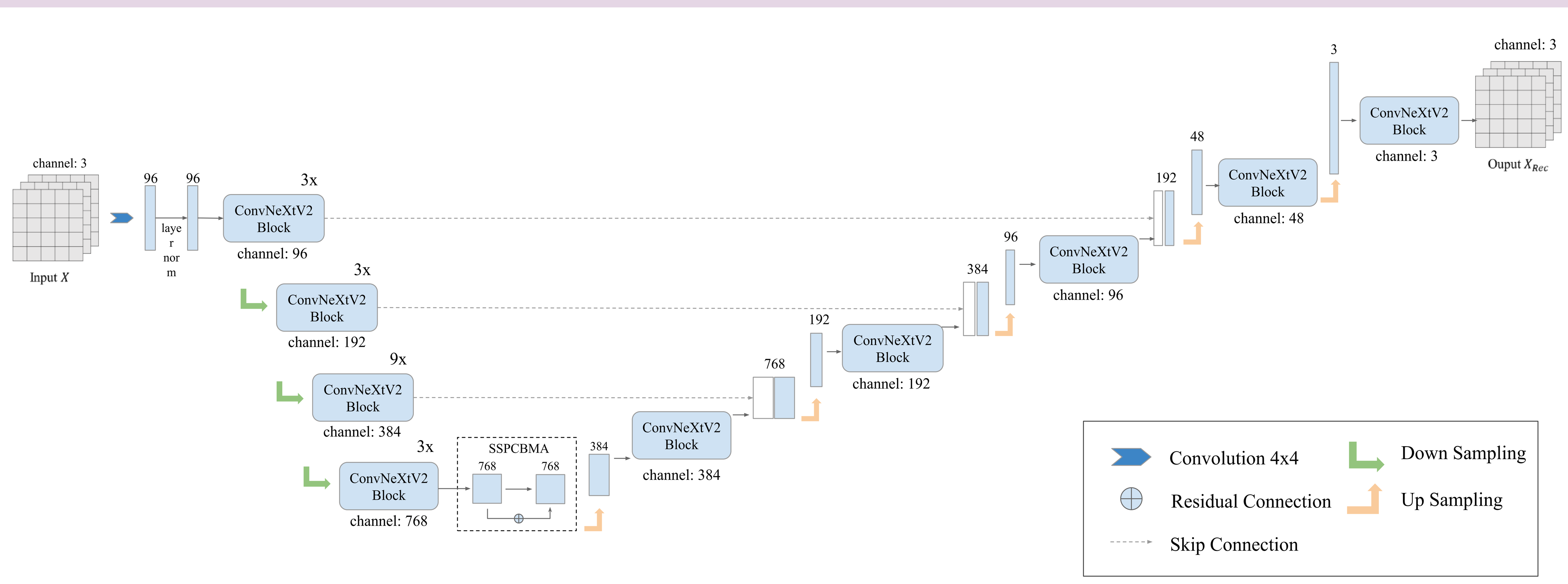
$$
\begin{aligned}
Multi\_Loss(t)_{S_i} = {} & \alpha_1(t)\cdot Loss_{L1smooth}(X_{Si}, X_{Ri}) \; + \\
& \alpha_2(t)\cdot Loss_{1-SSIM}(X_{Si}, X_{Ri}) \; + \\
& \alpha_3(t)\cdot Loss_{Focal}(PredictedMask_{S_i}, GT_{Si}) \; + \\
& \alpha_4(t)\cdot Loss_{Dice}(PredictedMask_{S_i}, GT_{Si})
\end{aligned}
$$

where $\alpha_1+\alpha_2+\alpha_3+\alpha_4 = 1.$

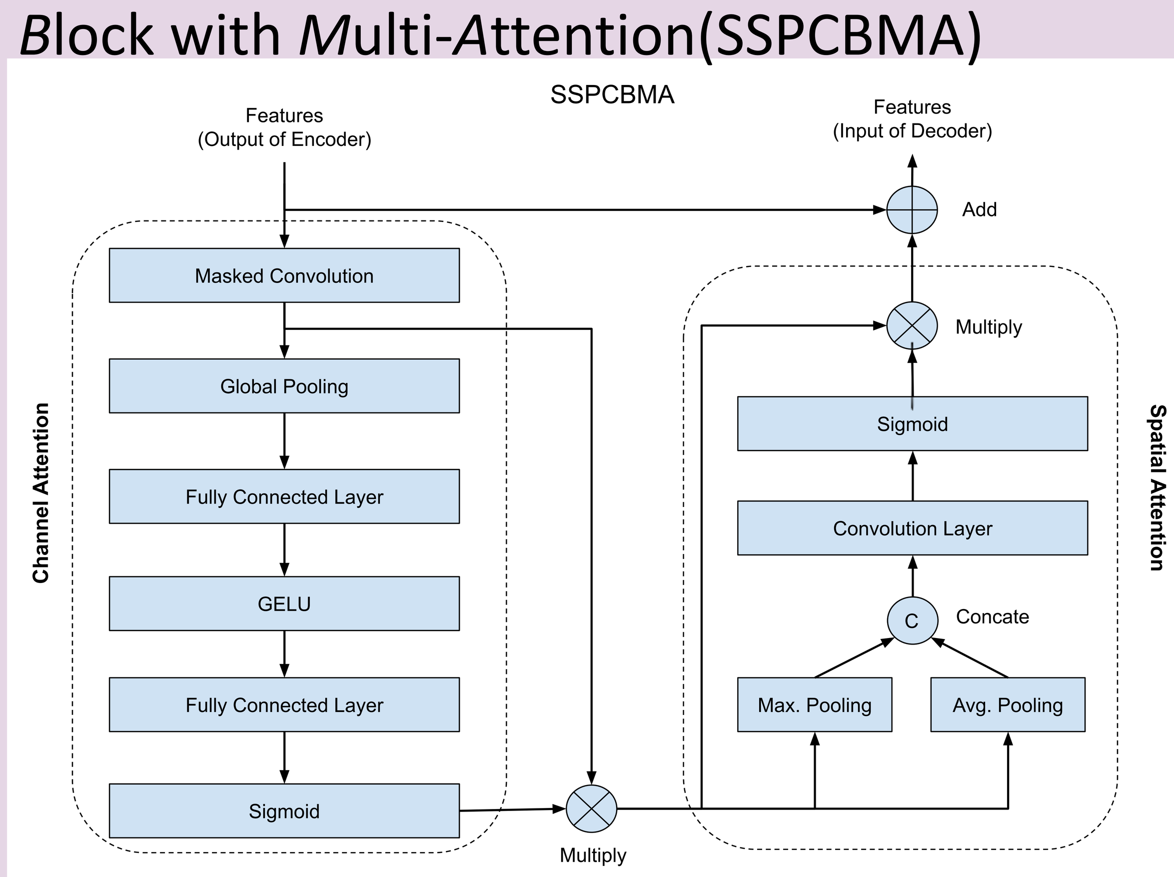
$$
\begin{aligned}
Synthesis\_Loss(t)_{Total} = {} & \beta_1(t)\cdot Multi\_Loss_{s_1} \; + \\
& \beta_2(t)\cdot Mutli\_Loss_{s_2} \; + \\
& \beta_3(t)\cdot Mutli\_Loss_{s_3} \; + \\
& \beta_4(t)\cdot Mutli\_Loss_{s_4}
\end{aligned}
$$

where $\beta_1+\beta_2+\beta_3+\beta_4 = 1.$

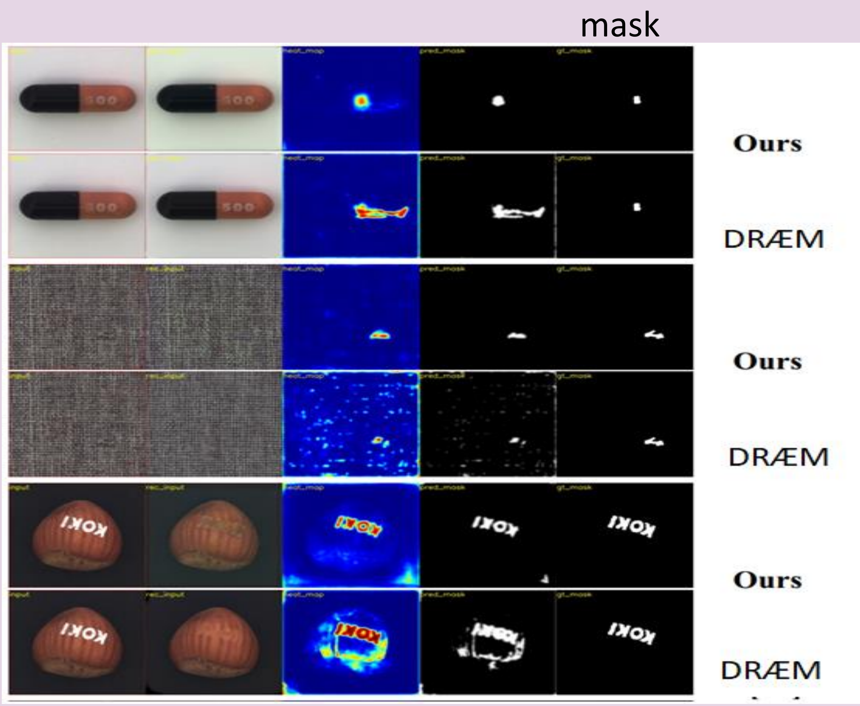
Reconstruction Unet(ConvNeXtUnetV2)



*S*elf-*S*upervised *P*redictive *C*onvolutional *B*lock with *M*ulti-*A*ttention(SSPCBMA)



Experiment Result



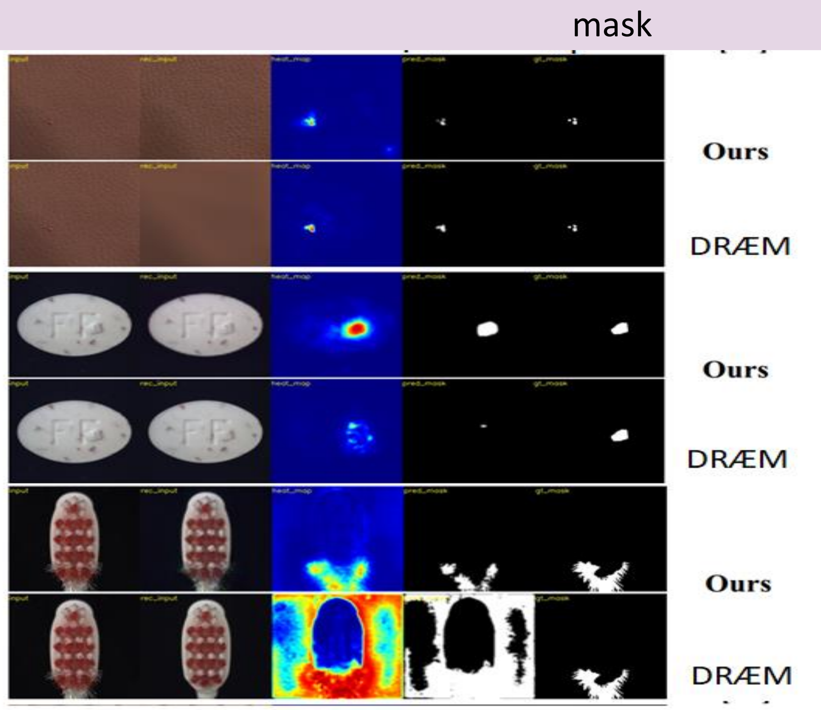


Image-level AUROC/ Pixel-level AUROC

Category	CutPaste	DRÆM	SSPCAB	RD	NSA	DSR	Patchcore	Ours
Carpet	93.9 / 98.3	96.9 / 97.5	93.1 / 92.6	98.7 / 98.9	95.6 / 95.5	**100.0** / 95.5	99.1 / 99.0	99.6 / **99.4**
Grid	**100.0** / 97.5	99.9 / 99.7	99.7 / 99.5	**100.0** / 98.3	99.9 / 99.2	**100.0** / 99.6	97.3 / 98.7	**100.0 / 99.8**
Leather	**100.0** / 99.5	**100.0** / 99.0	98.7 / 96.3	**100.0** / 99.4	99.9 / 99.5	**100.0 / 99.6**	**100.0** / 99.3	**100.0 / 99.6**
Tile	94.6 / 90.5	**100.0** / 99.2	**100.0** / 99.4	99.7 / 95.7	**100.0** / 99.3	**100.0** / 98.2	99.3 / 95.8	**100.0 / 99.5**
Wood	99.1 / 95.5	99.5 / 95.5	98.4 / **96.5**	99.5 / 95.8	97.5 / 90.7	96.3 / 92.5	**99.6** / 95.1	96.8 / 96.2
Average	97.5 / 96.3	99.3 / 98.2	98.0 / 96.9	**99.6** / 97.6	98.6 / 96.8	99.3 / 97.1	99.1 / 97.6	99.3 / **98.9**
Bottle	98.2 / 97.6	98.0 / 99.1	95.6 / **99.2**	**100.0** / 98.8	97.7 / 98.3	**100.0** / 98.9	**100.0** / 98.6	**100.0** / 98.4
Cable	81.2 / 90.0	90.9 / 95.2	92.7 / 95.1	96.1 / 97.2	94.5 / 96.0	93.8 / 96.7	**99.9 / 98.5**	95.7 / 94.4
Capsule	98.2 / 97.4	91.3 / 88.1	96.9 / 90.2	96.1 / 98.7	95.2 / 97.6	98.1 / 95.4	98.0 / 99.0	**99.0 / 99.1**
Hazelnut	98.3 / 97.3	**100.0 / 99.7**	**100.0 / 99.7**	**100.0** / 99.0	94.7 / 97.6	95.6 / 99.2	**100.0** / 98.7	98.9 / 99.5
Metal Nut	99.9 / 93.1	**100.0 / 99.6**	**100.0** / 99.4	**100.0** / 97.3	98.7 / 98.4	98.5 / 93.7	99.9 / 98.3	**100.0** / 98.9
Pill	94.9 / 95.7	97.1 / 97.3	97.4 / 97.2	98.7 / 98.1	**99.2 / 98.5**	97.5 / 93.4	97.5 / 97.6	97.2 / 97.3
Screw	88.7 / 96.7	**98.7** / 99.3	97.8 / 99.0	97.8 / **99.7**	90.2 / 96.5	96.2 / 98.5	98.2 / 99.5	97.5 / 99.3
Toothbrush	99.4 / 98.1	**100.0** / 97.3	97.9 / 97.3	**100.0** / 99.1	**100.0** / 94.9	99.7 / **99.5**	**100.0** / 98.6	99.2 / 99.3
Transistor	96.1 / 93.0	91.7 / 85.2	88.0 / 84.8	95.5 / 92.3	95.1 / 88.0	97.8 / 83.2	**99.9 / 96.5**	96.3 / 92.8
Zipper	99.9 / 99.3	**100.0** / 99.1	**100.0** / 98.4	97.9 / 98.3	99.8 / 94.2	**100.0** / 98.9	99.5 / 98.9	**100.0 / 99.4**
Average	95.5 / 95.8	96.8 / 96.0	96.6 / 96.0	98.2 / 97.9	96.5 / 96.0	97.7 / 95.7	**99.3 / 98.4**	98.4 / 97.8
TotalAverage	96.1 / 96.0	97.6 / 96.7	97.1 / 96.3	98.7 / 97.8	97.2 / 96.3	98.2 / 96.2	**99.2** / 98.1	98.7 / **98.2**

Per Region Overlap (PRO) score/ Pixel-level Average Precision (Pixel-AP)

Category	CutPaste	DRÆM	SSPCAB	RD	NSA	DSR	Patchcore	Ours
Carpet	50.4 / -	92.9 / 65.1	86.4 / 48.6	95.4 / 56.5	85.0 / -	- / 78.2	95.5 / 62.2	**99.8 / 80.6**
Grid	91.5 / -	98.3 / 62.8	98.0 / 57.9	94.2 / 15.8	96.8 / -	- / 68.0	94.0 / 24.5	**99.1 / 77.0**
Leather	83.7 / -	97.4 / 72.9	94.0 / 60.7	98.2 / 47.6	98.7 / -	- / 62.5	96.9 / 45.3	**99.2 / 68.3**
Tile	54.4 / -	98.2 / **95.2**	98.1 / 96.1	85.6 / 54.1	95.3 / -	- / 93.9	91.3 / 56.2	**98.3** / 94.3
Wood	64.0 / -	90.3 / 74.6	92.8 / **78.9**	91.4 / 48.3	85.3 / -	- / 68.4	87.1 / 49.3	**96.7** / 74.9
Average	68.8 / -	95.4 / 74.1	93.9 / 68.4	93.0 / 44.5	92.2 / -	- / 74.2	93.0 / 47.5	**98.6 / 79.0**
Bottle	91.2 / -	**96.8** / 88.9	96.3 / **89.4**	96.3 / 78.0	92.9 / -	- / 91.5	95.4 / 76.8	**96.8** / 86.5
Cable	59.8 / -	81.0 / 56.4	80.4 / 52.0	94.1 / 52.6	89.9 / -	- / **70.4**	**96.8** / 67.0	95.2 / 66.8
Capsule	83.5 / -	82.7 / 39.6	92.5 / 46.4	95.5 / 47.2	91.4 / -	- / 53.3	93.4 / 46.0	**95.9 / 57.6**
Hazelnut	81.3 / -	**98.5** / 92.6	98.2 / **93.4**	96.9 / 60.7	93.6 / -	- / 87.3	90.9 / 53.2	96.9 / 87.4
Metal Nut	54.4 / -	97.0 / **97.0**	**97.7** / 94.7	94.9 / 78.6	94.6 / -	- / 67.5	92.6 / 86.6	96.9 / 89.3
Pill	83.1 / -	88.4 / 47.6	89.6 / 48.3	**96.7 / 76.5**	96.0 / -	- / 65.7	94.5 / 75.7	93.6 / 68.8
Screw	72.6 / -	95.0 / **66.5**	95.2 / 61.7	98.5 / 52.1	90.1 / -	- / 52.5	97.5 / 34.7	**99.4** / 54.4
Toothbrush	88.1 / -	85.6 / 45.5	85.5 / 39.3	92.3 / 51.1	90.7 / -	- / **74.2**	**94.0** / 37.9	92.6 / 62.5
Transistor	68.5 / -	70.4 / 39.0	62.5 / 38.1	83.3 / 54.1	75.3 / -	- / 41.1	**92.3 / 66.9**	72.4 / 49.2
Zipper	84.9 / -	96.8 / 77.6	95.2 / 76.4	95.3 / 57.5	89.2 / -	- / 78.5	96.1 / 62.3	**98.9 / 83.0**
Average	76.7 / -	89.2 / 65.1	89.3 / 64.0	**94.4** / 60.8	90.4 / -	- / 68.2	**94.4** / 60.7	93.4 / **70.6**
TotalAverage	74.1 / -	91.3 / 68.1	90.8 / 65.5	93.9 / 55.4	91.0 / -	- / 70.2	93.9 / 56.3	**95.4 / 73.4**