Supplementary Materials for "Contextual Affinity Distillation for Image Anomaly Detection"

A. Structural and Logical Anomaly Localization Results on MVTec LOCO

We separately report the structural and logical anomaly localization performance of evaluated methods on the LOCO [1] dataset in Table 7. US [3] and GCAD [1] perform best on structural anomaly detection and logical anomaly detection, respectively. Both of them show a bias towards one kind of anomaly. In contrast, our proposed method shows good performance on both structural and logical anomaly detection, leading to the best overall performance at 0.730.

Table 7. Structural and logical anomaly localization results. The best scores are in bold and the second-best scores are with underlines.

Method	Structural anomalies	Logical anomalies	Mean
AE	0.296	0.460	0.378
VAE	0.305	0.459	0.382
MNAD [9]	0.412	0.266	0.339
VM	0.124	0.325	0.225
f-AnoGAN [11]	0.209	0.460	0.334
SPADE [4]	0.368	0.536	0.451
US [3]	**0.756**	0.497	0.626
RD [5]	0.739	0.474	0.607
PatchCore-25 [10]	0.705	0.541	0.623
GCAD [1]	0.692	**0.711**	<u>0.701</u>
DSKD	<u>0.754</u>	<u>0.707</u>	**0.730**

B. AU sPRO Scores with Different Integration Limits

Table 8 shows the structural and logical anomaly localization performance of the evaluated methods with different integration limits. For all methods, the AU sPRO scores show higher with larger integration limits. Our proposed method outperforms all the other methods with any integration limits.

C. Architecture of Global Context Condensing Block (GCCB)

The architecture of GCCB is shown in Table 9. Following the design philosophy of *bottleneck* block for ResNets [6], we use conv 1×1 layers for reducing and increasing channel dimensions to reduce the model complexity. Without using BatchNorm (BN) layers [7], we enable biases that act similar to them for convolution layers as modern CNN [8] architectures do. We experimentally found that using biases instead of the BN layers performs better. The output sizes are computed under the default setting where a pre-trained WideResNet50 [12] is used as the teacher T, and the images are resized to 256×256.

D. More Visualization Results on MVtec LOCO

We visualize more detection results for each of the five categories for positive samples and negative samples. Figures 7 to 11 show the logical anomaly detection results for each category. Figures 12 to 16 show the structural anomaly detection results. We visualize positive samples on the left and negative samples on the right. Note that the local student is identical to the RD [5] method.

E. More Visualization Results on modified MVtec AD

We also visualize more logical anomaly samples from the modified MVTec AD [2] in Fig. 17. Note that MVTec AD [2] has 15 categories, while only 3 out of them contain logical anomaly test images. We show logical anomaly images from the cable, capsule, and transistor categories. It should also be noted that MVTec AD [2] has 1,258 anomalous test images but only 37 of them are defined as logical anomaly samples, making it an unbalanced dataset for structural and logical anomaly detection.

Table 8. The AU sPRO scores for different integration limits L.

Method	L = 0.01	L = 0.05	L = 0.1	L = 0.3	L = 1.0
AE	0.166	0.378	0.499	0.699	0.882
VAE	0.162	0.382	0.506	0.705	0.884
MNAD [9]	0.176	0.339	0.447	0.643	0.853
VM	0.086	0.225	0.314	0.493	0.740
f-AnoGAN [11]	0.152	0.334	0.442	0.624	0.827
SPADE [4]	0.225	0.451	0.587	0.790	0.927
US [3]	0.402	0.626	0.717	0.836	0.937
RD [5]	0.410	0.607	0.695	0.830	0.943
PatchCore-25 [10]	0.395	0.623	0.724	0.872	0.960
GCAD [1]	0.462	0.701	0.787	0.891	0.962
DSKD	**0.513**	**0.730**	**0.809**	**0.910**	**0.970**

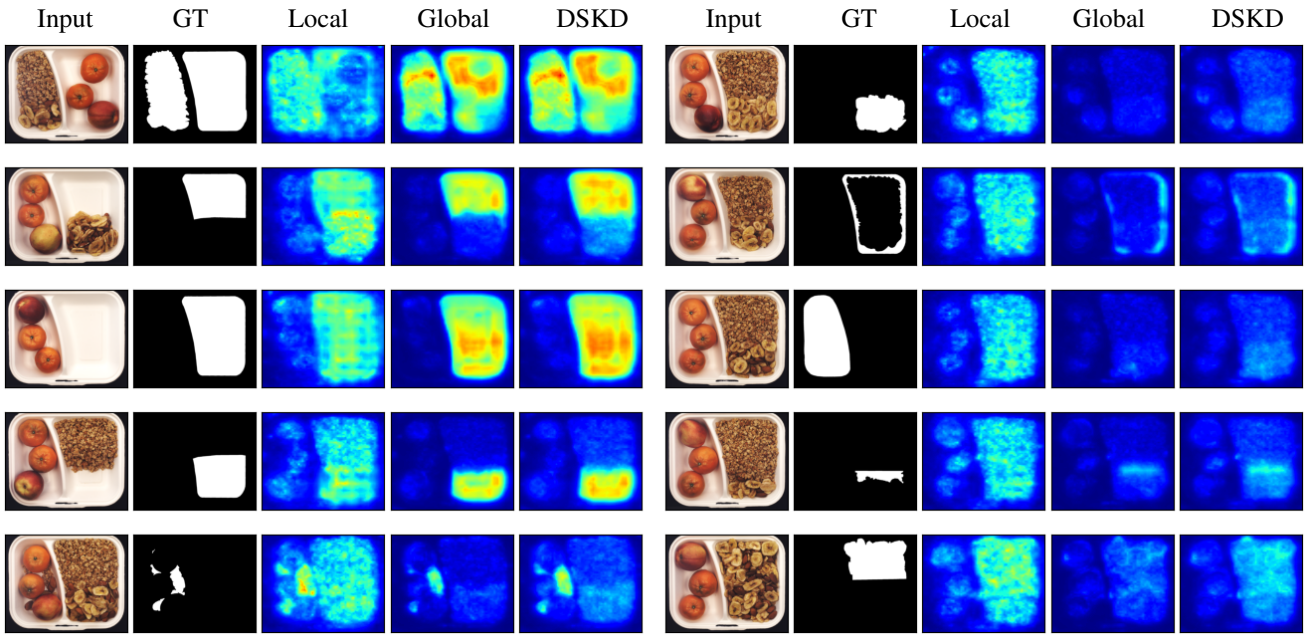

Figure 7. Logical anomaly samples for the Breakfast Box

Table 9. Architecture of GCCB.

Layer	Output Size	Kernel
Input	$8 \times 8 \times 2048$	
Conv1	$8 \times 8 \times 1024$	1×1
ReLU	$8 \times 8 \times 1024$	
Conv2	$1 \times 1 \times 1024$	8×8
ReLU	$1 \times 1 \times 1024$	
De_conv1	$8 \times 8 \times 1024$	8×8
ReLU	$8 \times 8 \times 1024$	
Conv3	$8 \times 8 \times 2048$	1×1
ReLU	$8 \times 8 \times 2048$	

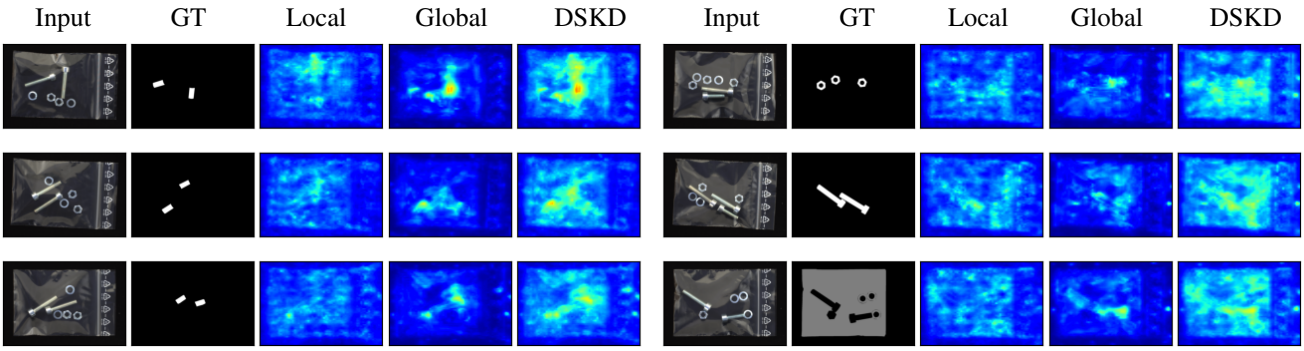

Figure 8. Logical anomaly samples for the Screw Bag

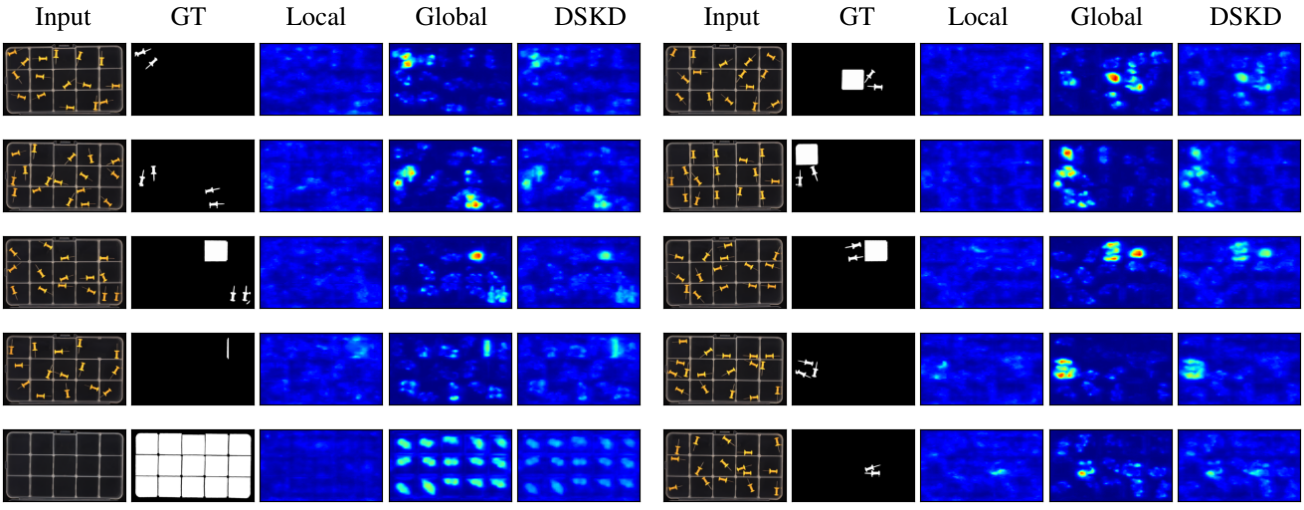

Figure 9. Logical anomaly samples for the Pushpins

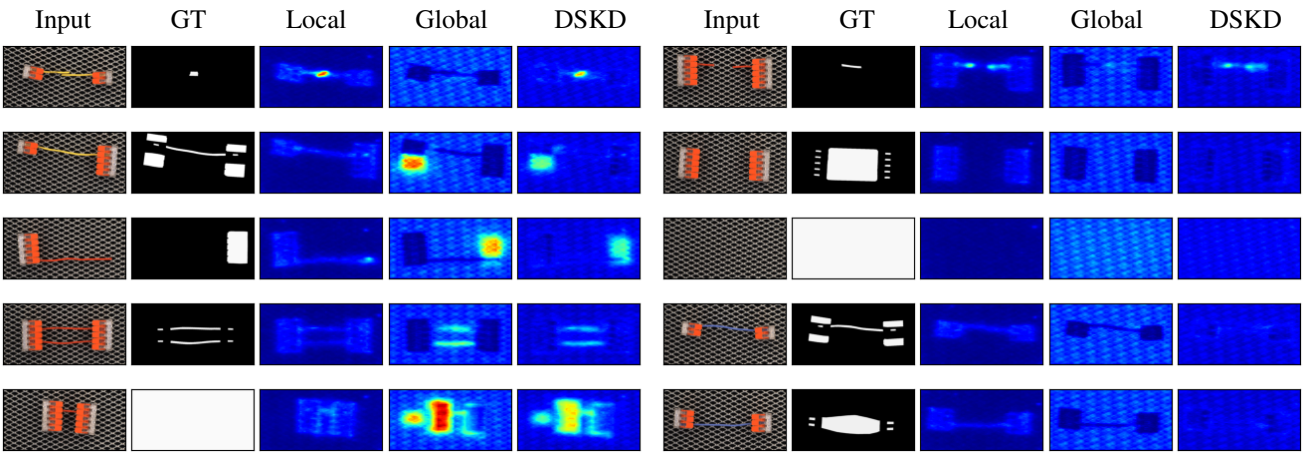

Figure 10. Logical anomaly samples for the Splicing Connectors

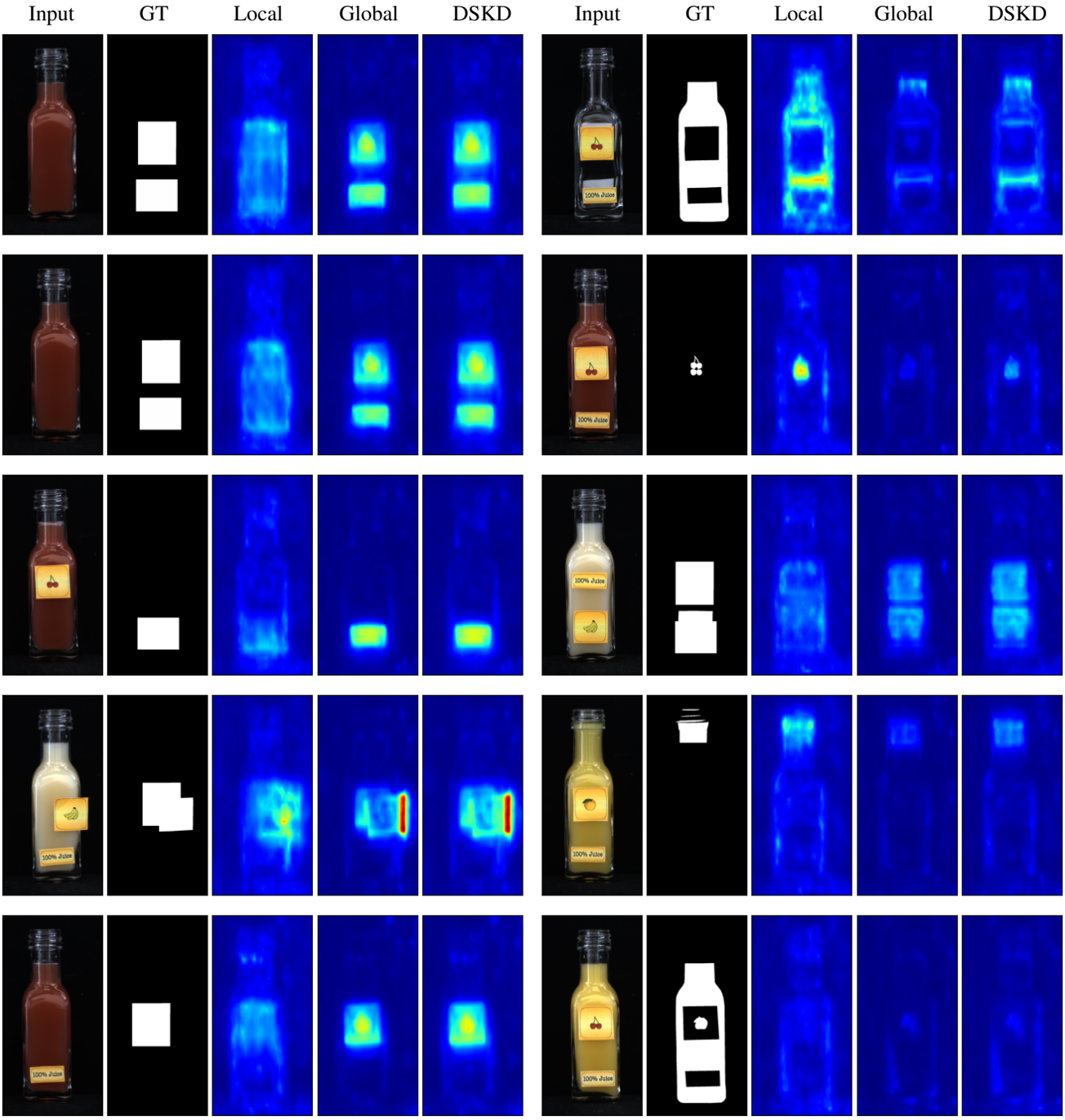

Figure 11. Logical anomaly samples for the Juice Bottle

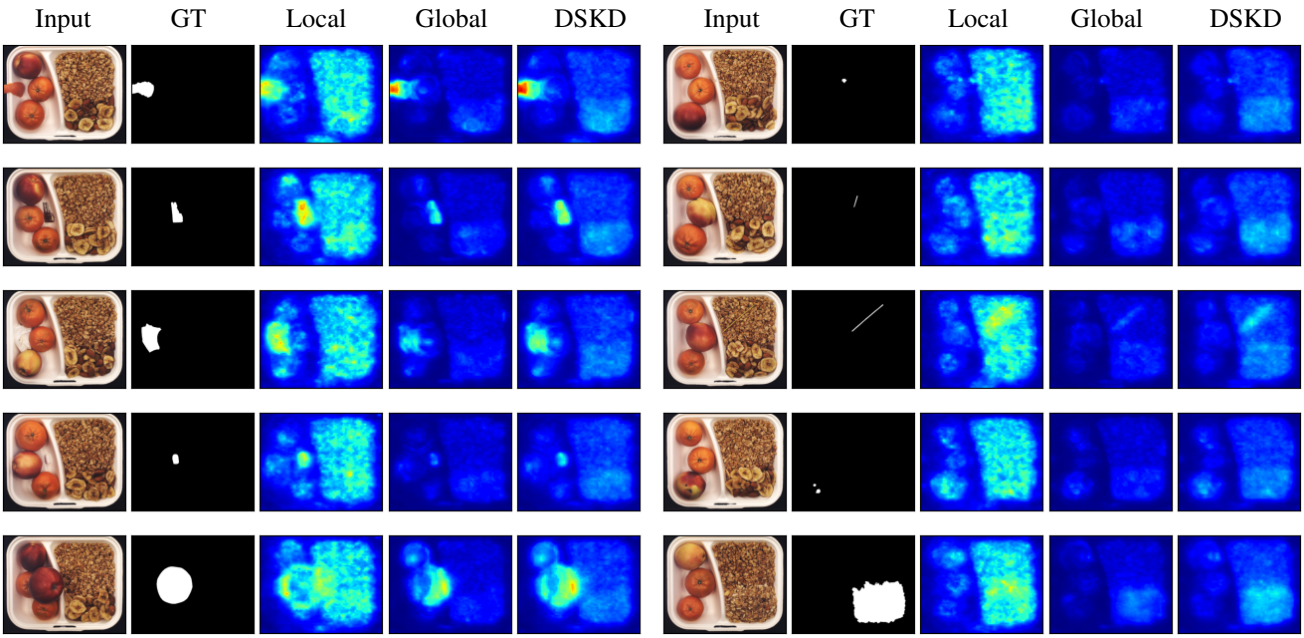

Figure 12. Structural anomaly samples for the Breakfast Box

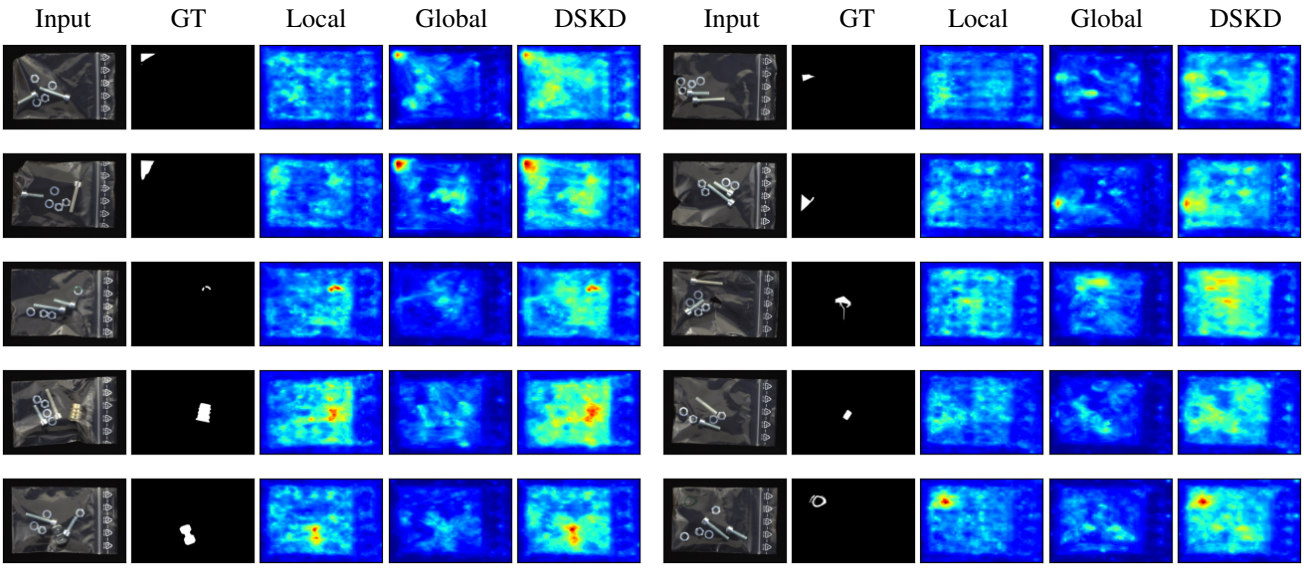

Figure 13. Structural anomaly samples for the Screw Bag

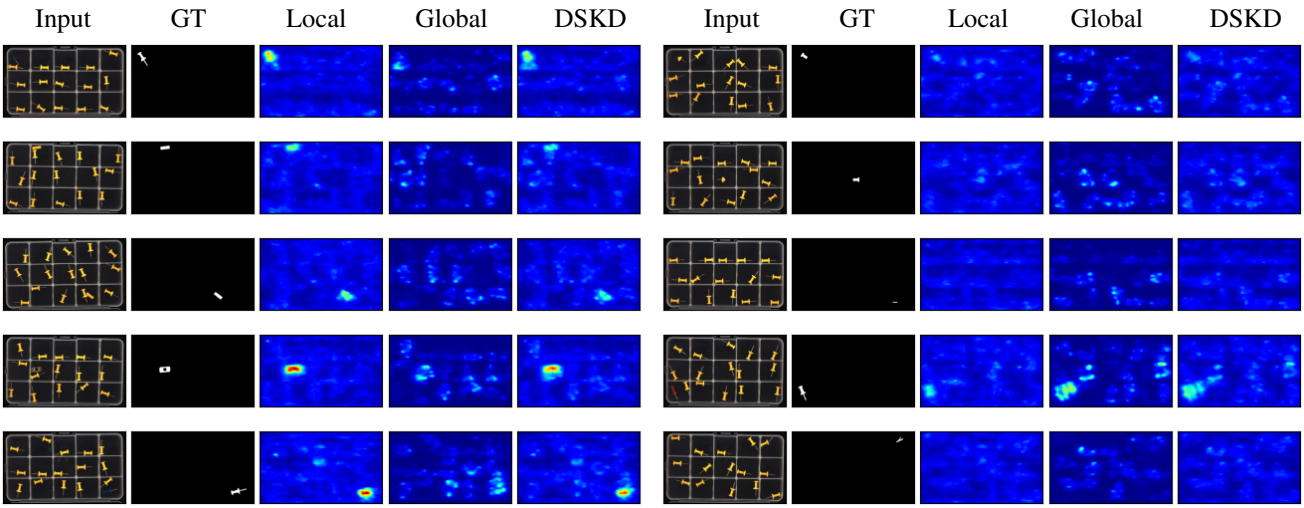

Figure 14. Structural anomaly samples for the Pushpins

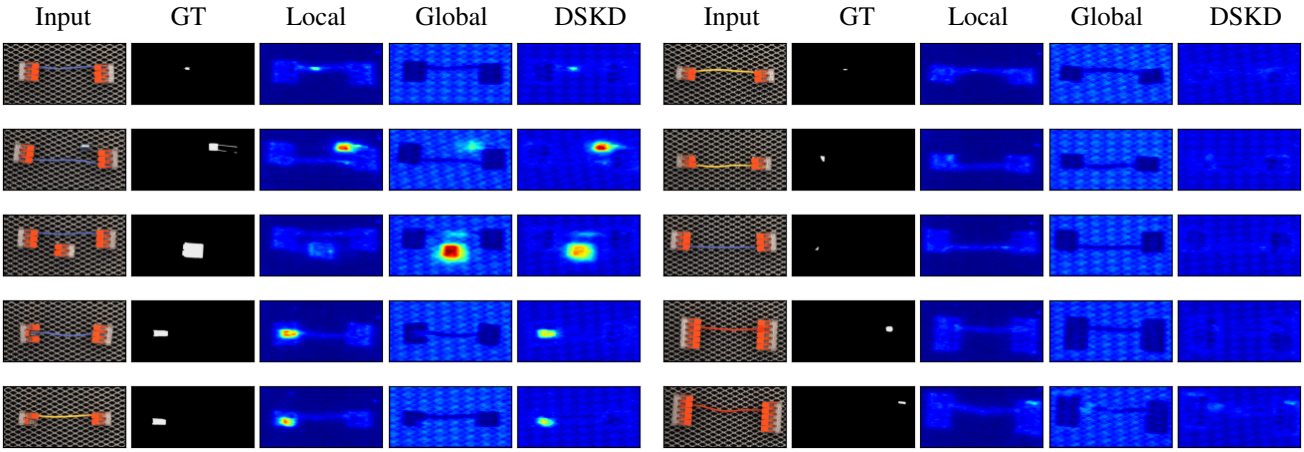

Figure 15. Structural anomaly samples for the Splicing Connectors

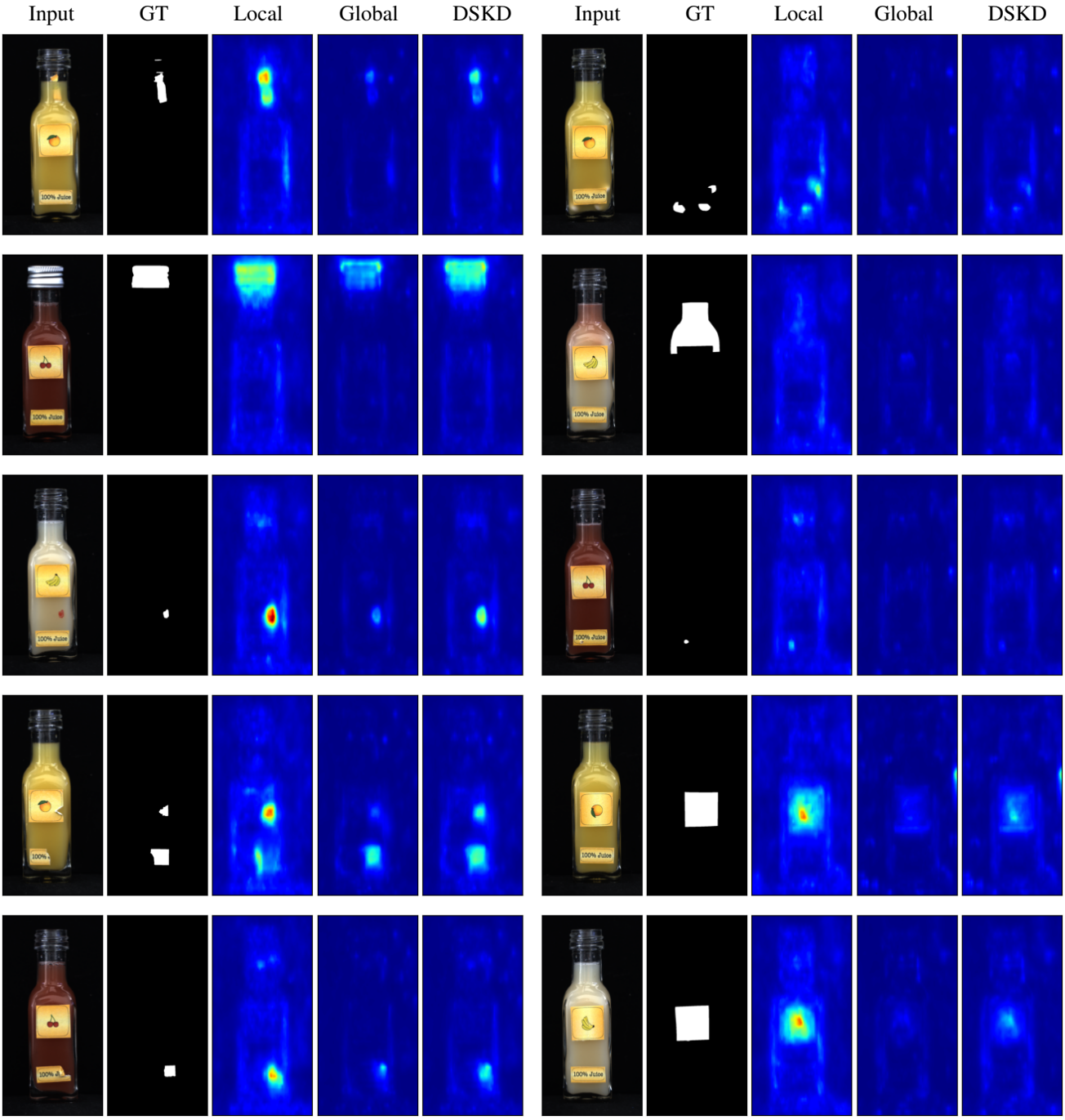

Figure 16. Structural anomaly samples for the Juice Bottle

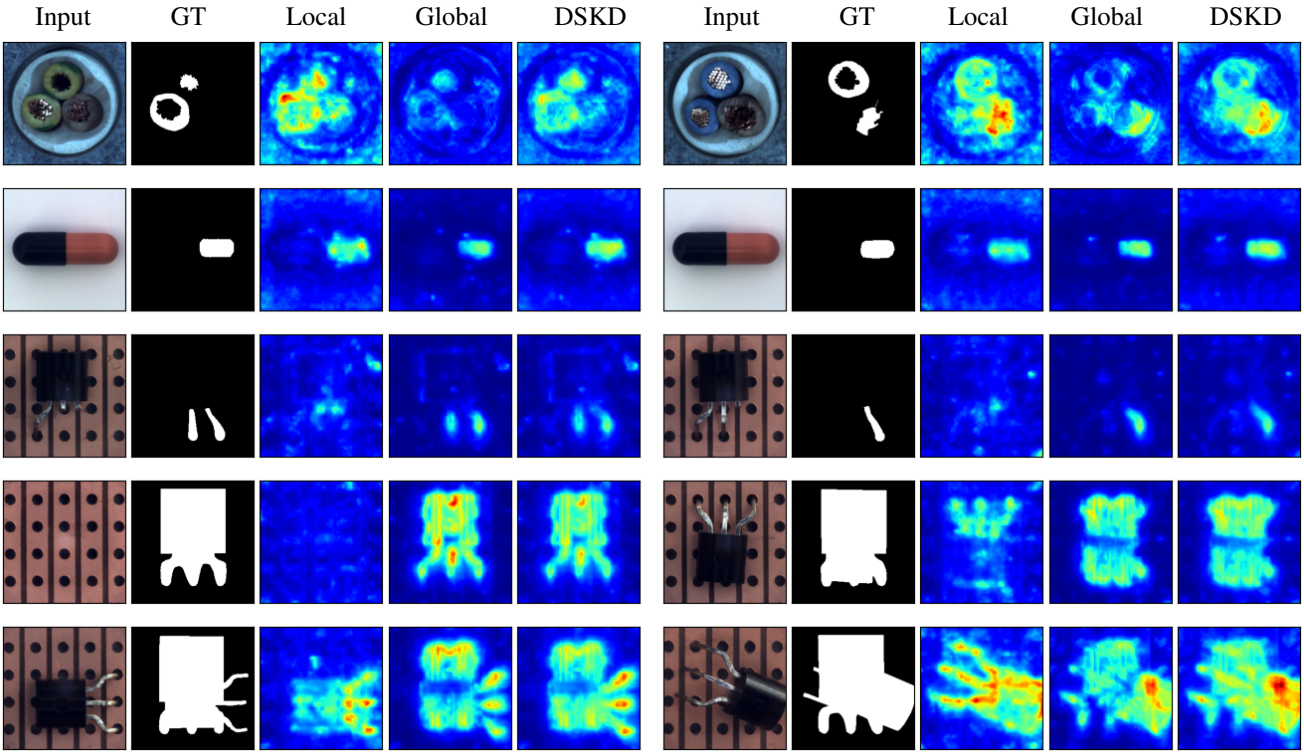

Figure 17. Logical anomaly samples from modified MVTec AD [2]

References

[1] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision*, 130(4):947–969, 2022. 1, 2

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 2, 9

[3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020. 1, 2

[4] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 1, 2

[5] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9737–9746, 2022. 1, 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 1

[8] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. 1

[9] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14372–14381, 2020. 1, 2

[10] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 1, 2

[11] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 1, 2

[12] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016. 1