Supplementary:
Anomaly Detection using Score-based Perturbation Resilience

1. Derivation of Eq. (2)

Let perturbation kernel $p(\tilde{\mathrm{x}}|\mathrm{x})$ be a gaussian kernel $\mathrm{N}(\tilde{\mathrm{x}};\mathrm{x},\sigma^2 I) = C\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{x}\|^2)$, then

$$\nabla_{\tilde{\mathrm{x}}}\log p(\tilde{\mathrm{x}}) = \frac{\nabla_{\tilde{\mathrm{x}}}p(\tilde{\mathrm{x}})}{p(\tilde{\mathrm{x}})} \tag{1}$$

$$= \frac{\nabla_{\tilde{\mathrm{x}}}\int p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})\mathrm{dx}}{p(\tilde{\mathrm{x}})} \tag{2}$$

$$= \int \frac{\nabla_{\tilde{\mathrm{x}}}p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})}{p(\tilde{\mathrm{x}})}\mathrm{dx}. \tag{3}$$

Considering $p(\tilde{\mathrm{x}}|\mathrm{x}) = C\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{x}\|^2)$, we have $\nabla_{\tilde{\mathrm{x}}}p(\tilde{\mathrm{x}}|\mathrm{x}) = \frac{(\mathrm{x}-\tilde{\mathrm{x}})}{\sigma^2}p(\tilde{\mathrm{x}}|\mathrm{x})$. Plug this in (3), we can rewrite as

$$\int \frac{(\mathrm{x}-\tilde{\mathrm{x}})}{\sigma^2}\frac{p(\mathrm{x})p(\tilde{\mathrm{x}}|\mathrm{x})}{p(\tilde{\mathrm{x}})}\mathrm{dx}. \tag{4}$$

By substituting $\frac{p(\mathrm{x})p(\tilde{\mathrm{x}}|\mathrm{x})}{p(\tilde{\mathrm{x}})}$ with $p(\mathrm{x}|\tilde{\mathrm{x}})$ according to Bayes' rule, we reform (4) as

$$\int \frac{(\mathrm{x}-\tilde{\mathrm{x}})}{\sigma^2}p(\mathrm{x}|\tilde{\mathrm{x}})\mathrm{dx} = \frac{1}{\sigma^2}\big(\mathrm{E}[\mathrm{x}|\tilde{\mathrm{x}}]-\tilde{\mathrm{x}}\big). \tag{5}$$

2. Derivation of Eq. (3)

We assumed that $p(\mathrm{x})$ is the uniform distribution on the data manifold $\mathsf{M} = \mathrm{supp}(\mathrm{X})$ and M satisfies the local linearity. Then, the conditional expectation in Eq. (3) can be reformulated as

$$\mathrm{E}[\mathrm{x}|\tilde{\mathrm{x}}] = \int_{\mathrm{x}\in\mathsf{M}} \mathrm{x}p(\mathrm{x}|\tilde{\mathrm{x}})\mathrm{dx} \tag{6}$$

$$= \int \mathrm{x}\frac{p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})}{p(\tilde{\mathrm{x}})}\mathrm{dx} \tag{7}$$

$$= \int \mathrm{x}\frac{p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})}{\int p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})\mathrm{dx}}\mathrm{dx}, \tag{8}$$

where $p(\tilde{\mathrm{x}}) = \int p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})\mathrm{dx}$. Since $p(\mathrm{x})$ is the uniform distribution, we can regard it as constant and eliminate it in the denominator and numerator. In addition, considering $p(\tilde{\mathrm{x}}|\mathrm{x})$ is a gaussian kernel,

$$\int \mathrm{x}\frac{p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})}{\int p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})\mathrm{dx}}\mathrm{dx} = \int \mathrm{x}\frac{C\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{x}\|^2)}{\int C\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{x}\|^2)\mathrm{dx}}\mathrm{dx}. \tag{9}$$

Let $\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})$ denote a projected point onto M from $\tilde{\mathrm{x}}$, then $\|\tilde{\mathrm{x}}-\mathrm{x}\|^2 = \|\tilde{\mathrm{x}}-\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2 + \|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})-\mathrm{x}\|^2$ by Pythagoras' theorem. Thus, (9) can be rewritten as

$$= \int \mathrm{x}\frac{\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2 - \frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})-\mathrm{x}\|^2)}{\int \exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2 - \frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})-\mathrm{x}\|^2)\mathrm{dx}}\mathrm{dx} \tag{10}$$

$$= \int \mathrm{x}\frac{\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2)\exp(-\frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})-\mathrm{x}\|^2)}{\int \exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}}-\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2)\exp(-\frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})-\mathrm{x}\|^2)\mathrm{dx}}\mathrm{dx}. \tag{11}$$

Finally, since $\exp(-\frac{1}{2\sigma^2}\|\tilde{\mathrm{x}} - \mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}})\|^2)$ is independent to x, we can derive as

$$\int \mathrm{x} \frac{\exp(-\frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}}) - \mathrm{x}\|^2)}{\int \exp(-\frac{1}{2\sigma^2}\|\mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}}) - \mathrm{x}\|^2)\mathrm{dx}} \mathrm{dx} = \mathrm{P}_{\mathsf{M}}(\tilde{\mathrm{x}}). \tag{12}$$

3. Network architecture

Layers	Filter	Channels	Stride	Input	Output	Resolution
[Conv2D, GroupNorm, Swish] × 2	3 × 3	64	1	*Input*	E_{11}	$H \times W$
TimeEmbedding	-	-	-	E_{11}	E_{12}	$H \times W$
MaxPooling	2 × 2	-	2	E_{12}	E_{21}	$H/2 \times W/2$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	128	1	E_{21}	E_{22}	$H/2 \times W/2$
TimeEmbedding	-	-	-	E_{22}	E_{23}	$H/2 \times W/2$
MaxPooling	2 × 2	-	2	E_{23}	E_{31}	$H/4 \times W/4$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	256	1	E_{31}	E_{32}	$H/4 \times W/4$
TimeEmbedding	-	-	-	E_{32}	E_{33}	$H/4 \times W/4$
MaxPooling	2 × 2	-	2	E_{33}	E_{41}	$H/8 \times W/8$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	512	1	E_{41}	E_{42}	$H/8 \times W/8$
TimeEmbedding	-	-	-	E_{42}	E_{43}	$H/8 \times W/8$
MaxPooling	2 × 2	-	2	E_{43}	E_{51}	$H/16 \times W/16$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	512	1	E_{51}	E_{52}	$H/16 \times W/16$
TimeEmbedding	-	-	-	E_{52}	E_{53}	$H/16 \times W/16$
Upsampling	-	-	-	E_{53}	E_{54}	$H/8 \times W/8$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	256	1	$Concat[E_{54}, E_{43}]$	D_{11}	$H/8 \times W/8$
TimeEmbedding	-	-	-	D_{11}	D_{12}	$H/8 \times W/8$
Upsampling	-	-	-	D_{12}	D_{13}	$H/4 \times W/4$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	128	1	$Concat[D_{13}, E_{33}]$	D_{21}	$H/4 \times W/4$
TimeEmbedding	-	-	-	D_{21}	D_{22}	$H/4 \times W/4$
Upsampling	-	-	-	D_{22}	D_{23}	$H/2 \times W/2$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	64	1	$Concat[D_{23}, E_{23}]$	D_{31}	$H/2 \times W/2$
TimeEmbedding	-	-	-	D_{31}	D_{32}	$H/2 \times W/2$
Upsampling	-	-	-	D_{32}	D_{32}	$H \times W$
[Conv2D, GroupNorm, Swish] × 2	3 × 3	64	1	$Concat[D_{33}, E_{13}]$	D_{41}	$H \times W$
TimeEmbedding	-	-	-	D_{41}	D_{42}	$H \times W$
Conv2D	1 × 1	3	1	D_{42}	*Output*	$H \times W$

Table 1. The U-net-based network architecture of our method.

Tab. 1 shows the U-net-based network architecture utilized in this paper [2]. The network includes Group Normalization [4] with a group number of 32, a swish activation function [1] for non-linearity, and a sinusoidal time embedding used in [3] with a dimension of 256.

4. Examples of anomaly localization

MVTec AD From Fig. 1 to Fig. 15, we show examples of anomaly localization for each categories in MVTec AD. The categories include: Carpet (Fig. 1), Grid (Fig. 2), Leather (Fig. 3), Tile (Fig. 4), Wood (Fig. 5), Bottle

(Fig. 6), Cable (Fig. 7), Capsule (Fig. 8), Hazelnut (Fig. 9), Metal nut (Fig. 10), Pill (Fig. 11), Screw (Fig. 12), Tooth brush (Fig. 13), Transistor (Fig. 14), and Zipper (Fig. 15).

BTAD For each category in BTAD, from Fig. 16 to Fig. 18, we present some examples of anomaly localization. The categories include: Product 01 (Fig. 16), Product 02 (Fig. 17), and Product 03 (Fig. 18)

MPDD We provide some examples of anomaly localization for each category in MPDD from Fig. 19 to Fig. 24. The categories include: Bracket black (Fig. 19), Bracket brown (Fig. 20), Bracket white (Fig. 21), Connector (Fig. 22), Metal Plate (Fig. 23), and Tubes (Fig. 24)

References

[1] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017. 2

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[3] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2

[4] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 2

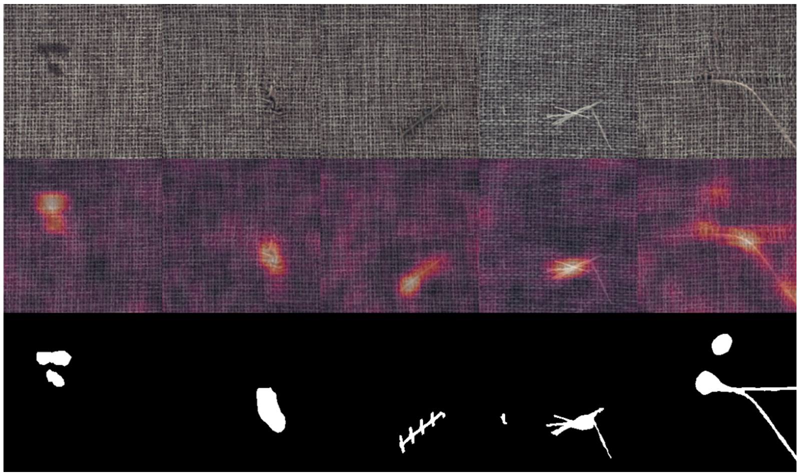

Figure 1. Anomaly localization examples of the carpet category. From top to bottom: input image, restoration error map, ground truth.

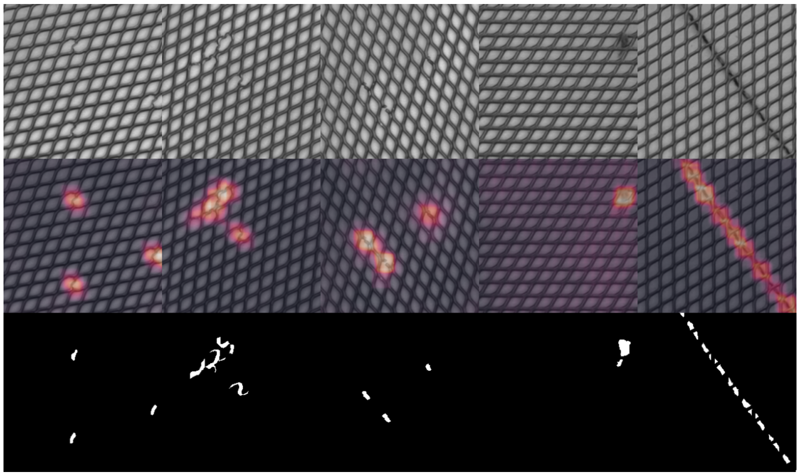

Figure 2. Anomaly localization examples of the grid category. From top to bottom: input image, restoration error map, ground truth.

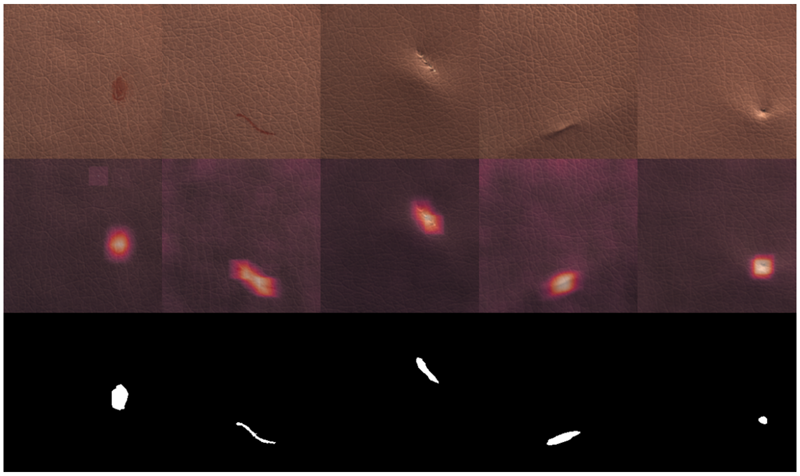

Figure 3. Anomaly localization examples of the leather category. From top to bottom: input image, restoration error map, ground truth.

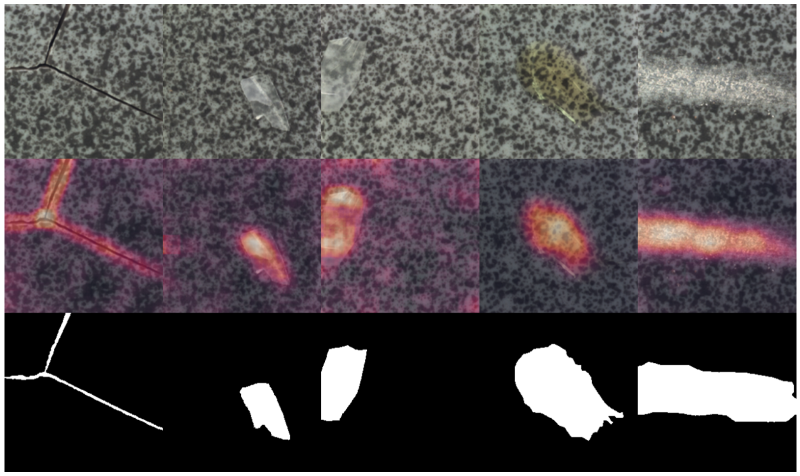

Figure 4. Anomaly localization examples of the tile category. From top to bottom: input image, restoration error map, ground truth.

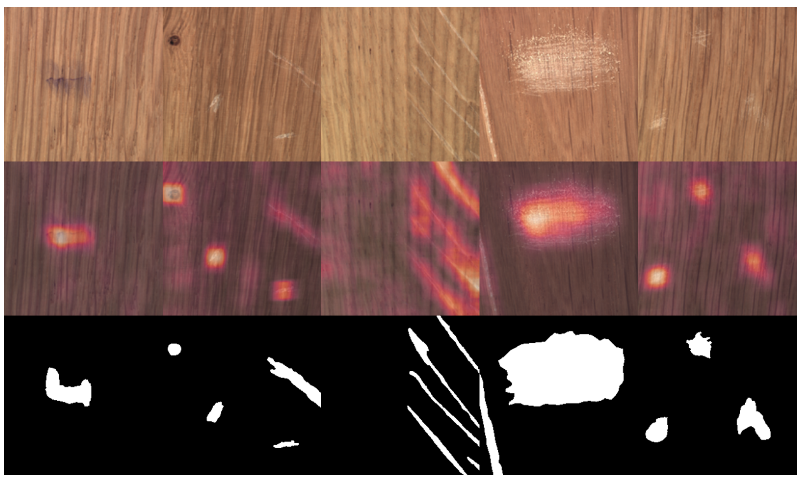

Figure 5. Anomaly localization examples of the wood category. From top to bottom: input image, restoration error map, ground truth.

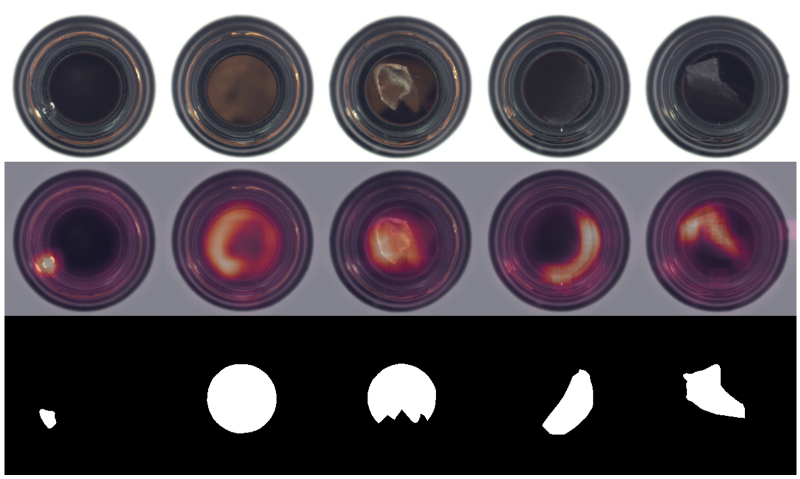

Figure 6. Anomaly localization examples of the bottle category. From top to bottom: input image, restoration error map, ground truth.

Figure 7. Anomaly localization examples of the cable category. From top to bottom: input image, restoration error map, ground truth.

Figure 8. Anomaly localization examples of the capsule category. From top to bottom: input image, restoration error map, ground truth.

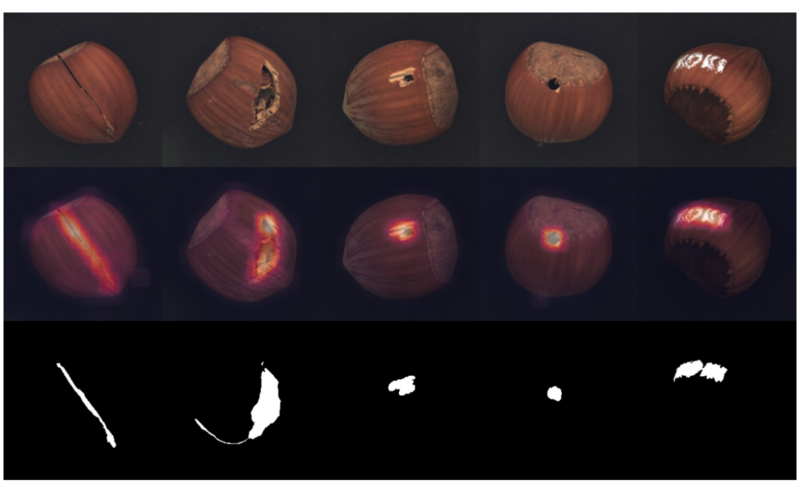

Figure 9. Anomaly localization examples of the hazelnut category. From top to bottom: input image, restoration error map, ground truth.

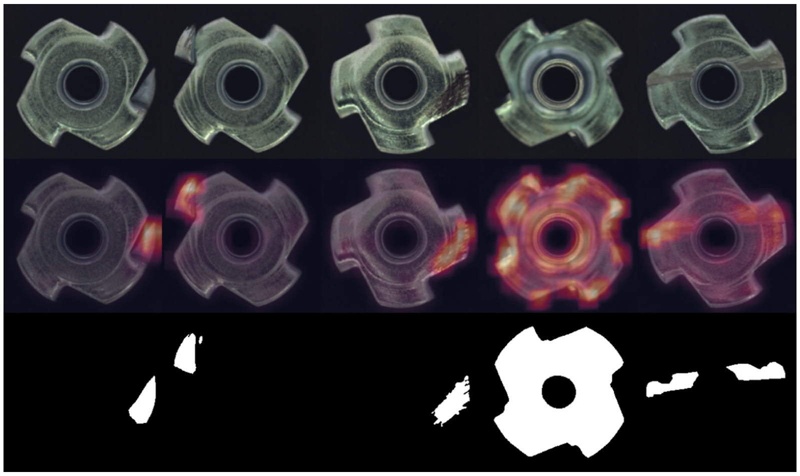

Figure 10. Anomaly localization examples of the metal nut category. From top to bottom: input image, restoration error map, ground truth.

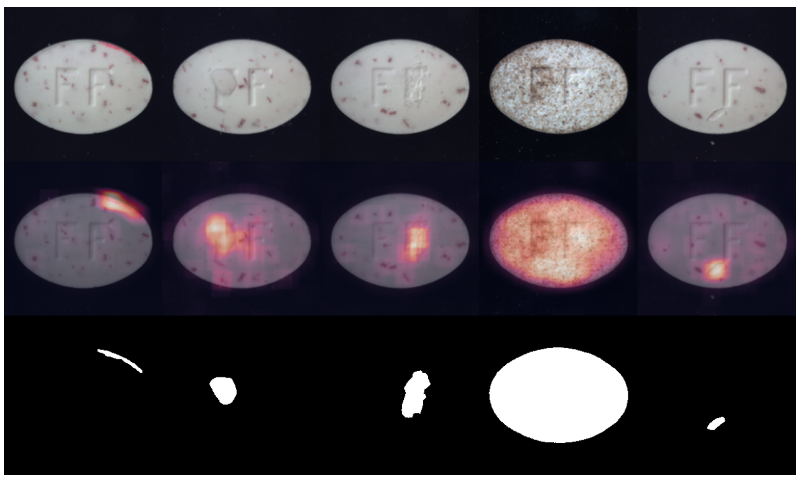

Figure 11. Anomaly localization examples of the pill category. From top to bottom: input image, restoration error map, ground truth.

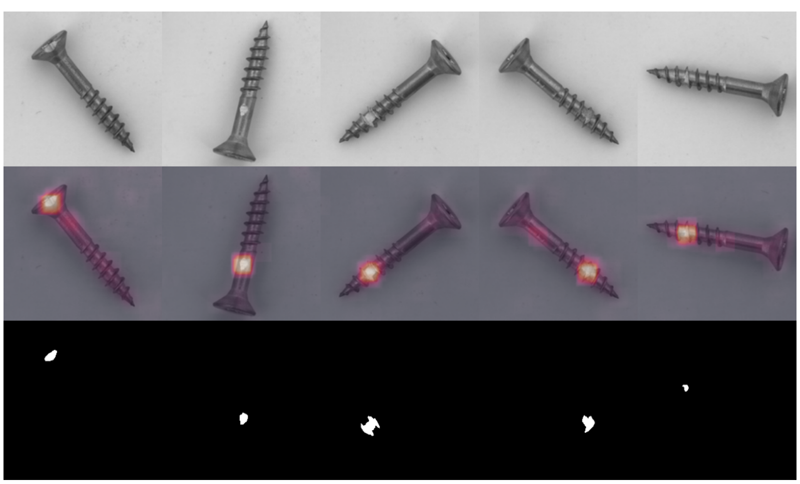

Figure 12. An anomaly localization examples of the screw category. From top to bottom: input image, restoration error map, ground truth.

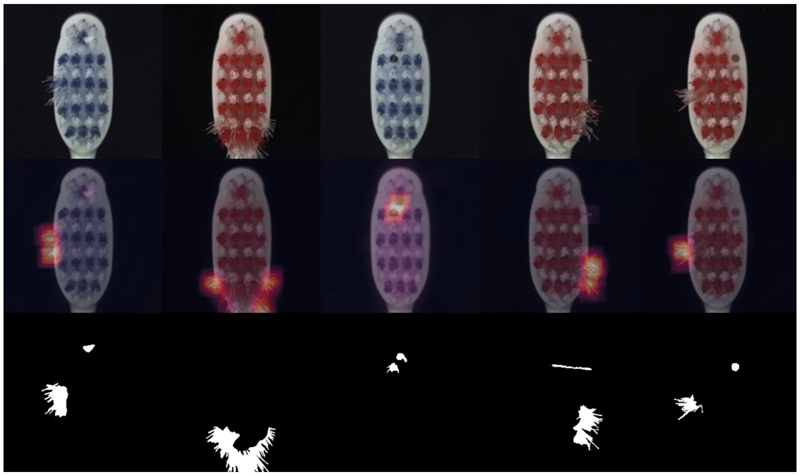

Figure 13. An anomaly localization examples of the toothbrush category. From top to bottom: input image, restoration error map, ground truth.

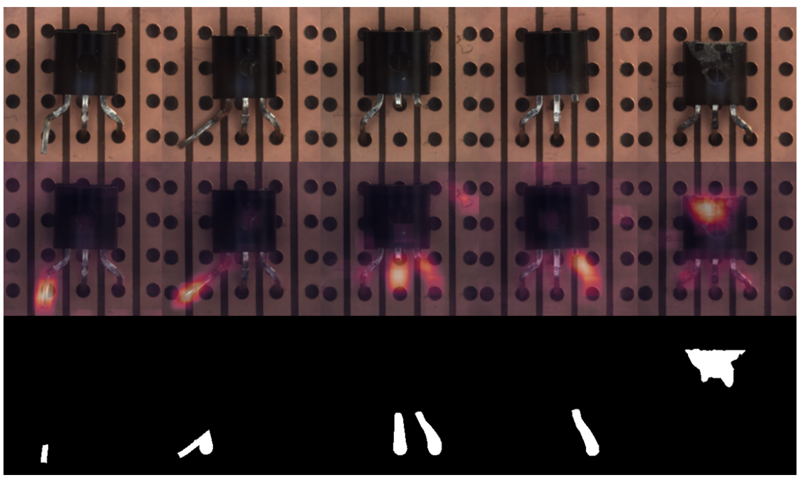

Figure 14. An anomaly localization examples of the transistor category. From top to bottom: input image, restoration error map, ground truth.

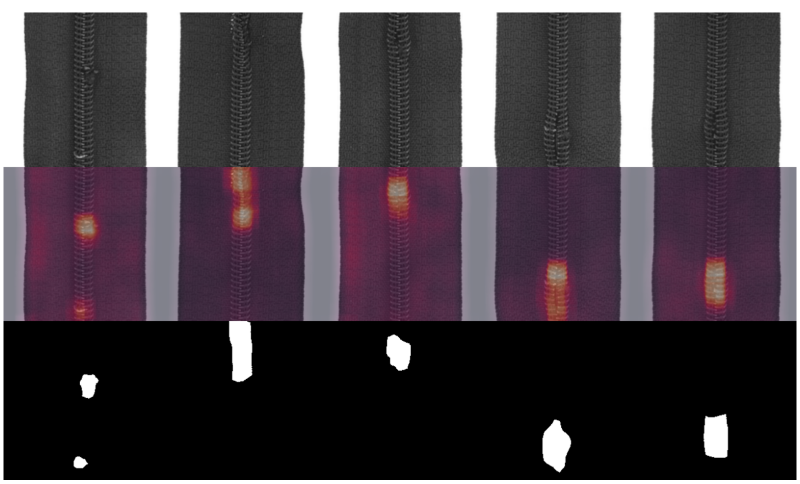

Figure 15. Anomaly localization examples of the zipper category. From top to bottom: input image, restoration error map, ground truth.

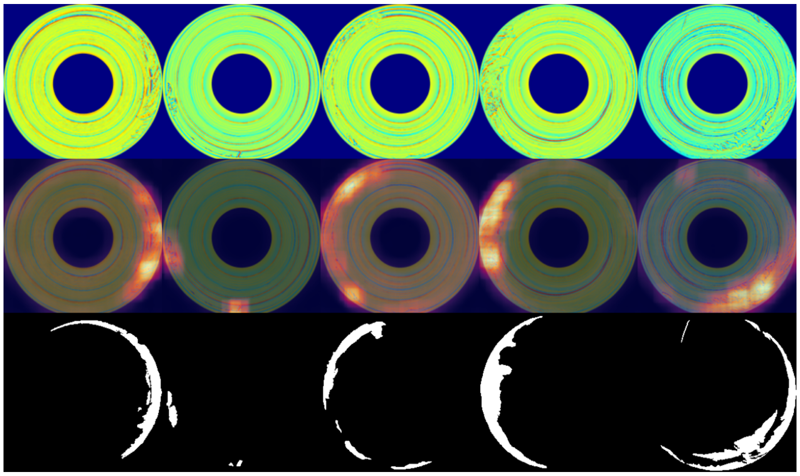

Figure 16. Anomaly localization examples of the product 01 category. From top to bottom: input image, restoration error map, ground truth.

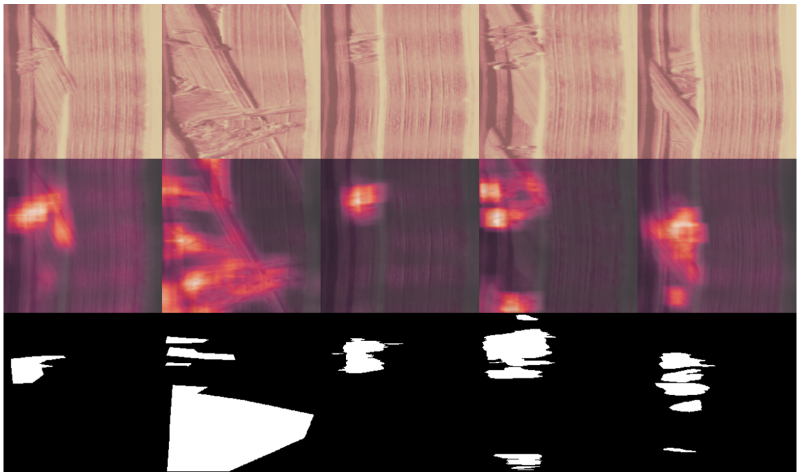

Figure 17. Anomaly localization examples of the product 02 category. From top to bottom: input image, restoration error map, ground truth.

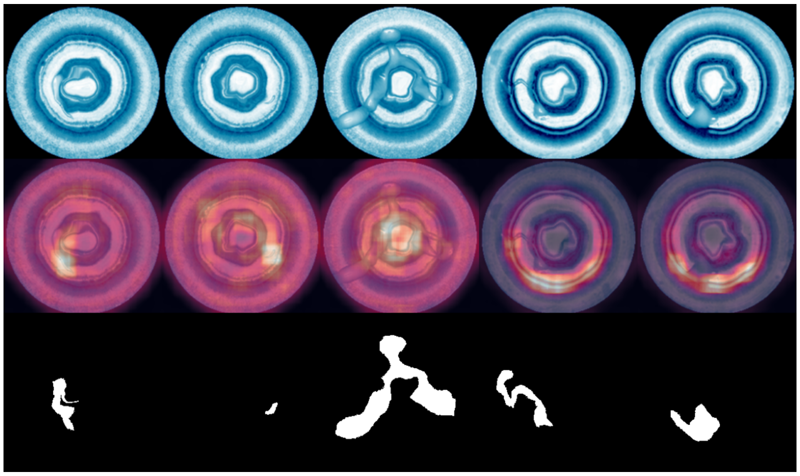

Figure 18. Anomaly localization examples of the product 03 category. From top to bottom: input image, restoration error map, ground truth.

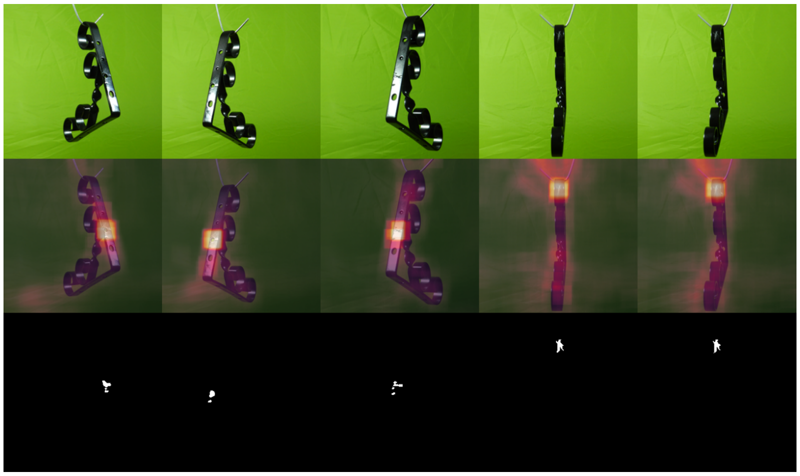

Figure 19. Anomaly localization examples of the bracket black category. From top to bottom: input image, restoration error map, ground truth.

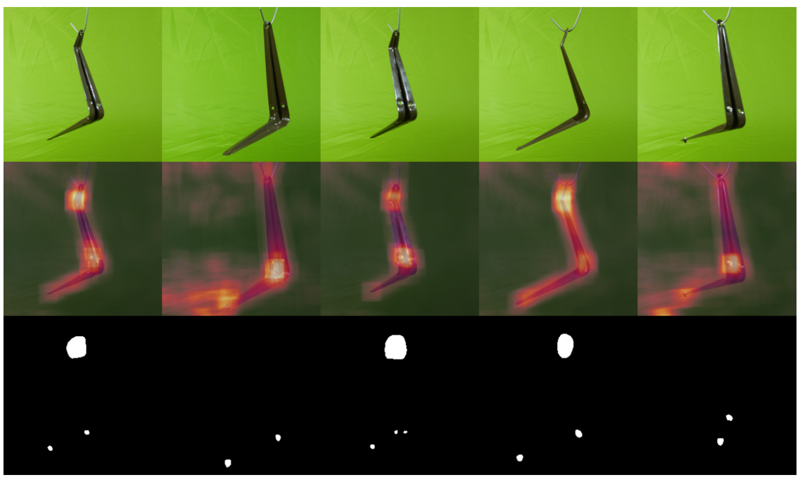

Figure 20. Anomaly localization examples of the bracket brown category. From top to bottom: input image, restoration error map, ground truth.

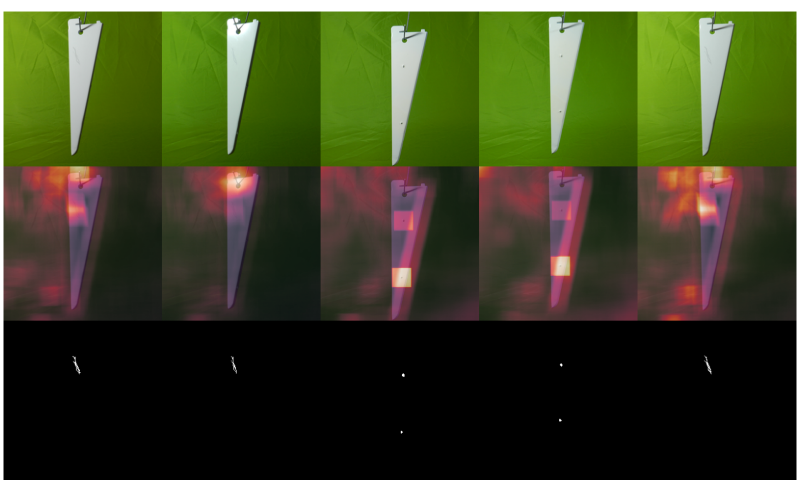

Figure 21. Anomaly localization examples of the bracket white category. From top to bottom: input image, restoration error map, ground truth.

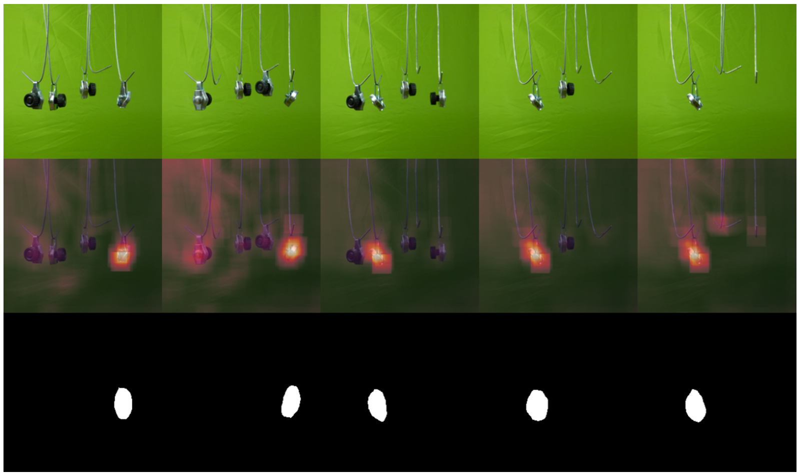

Figure 22. Anomaly localization examples of the connector category. From top to bottom: input image, restoration error map, ground truth.

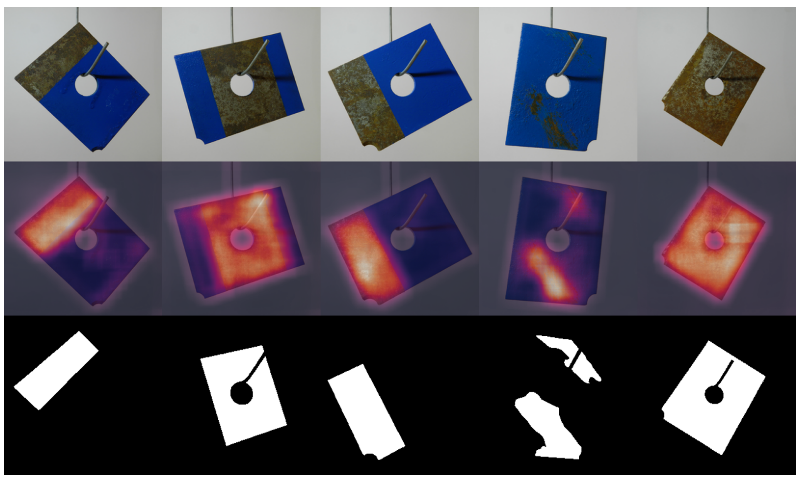

Figure 23. Anomaly localization examples of the metal plate category. From top to bottom: input image, restoration error map, ground truth.

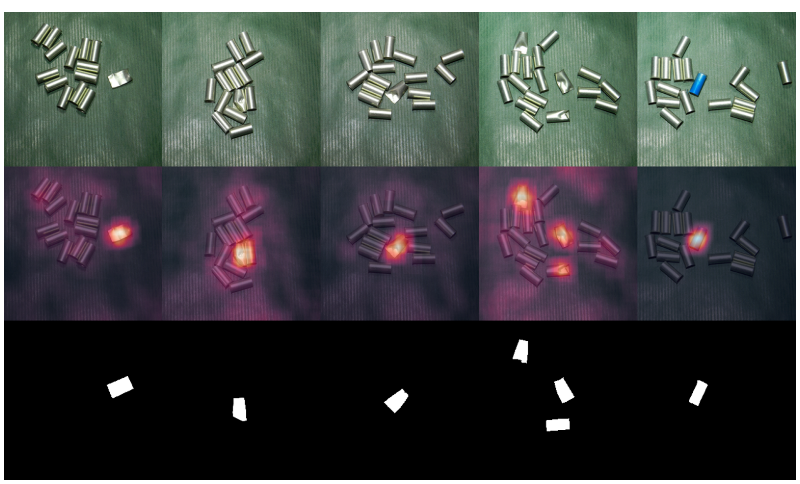

Figure 24. Anomaly localization examples of the tube category. From top to bottom: input image, restoration error map, ground truth.