Supplementary Material - GLAD: A Global-to-Local Anomaly Detector

Aitor Artola[1,2], Yannis Kolodziej[2], Jean-Michel Morel[1], Thibaud Ehret[1]

[1]Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, France
[2]Visionairy
aitor.artola@ens-paris-saclay.fr

1. Global weight decay

Figure 1 shows the decay of the weights π_k of the global model trained on the features of layers 2 and 3 of WideResnet50-2 for all classes of the MVTec dataset.

2. Quantile scoring

The figure 2 shows the evolution of the average AUROC on MvTec database with the features of WideResnet50-2 as a function of the quantile chosen for the image scoring.

3. Local Weight maps

We show different examples of weight and local sparsity maps learned on features from layer 1 of WideResnet50-2 in Figure 3 (bottle), Figure 4 (cable), Figure 5 (metal nut), Figure 6 (screw) and Figure 7 (zipper). We also show different examples of weight and local sparsity map on textures in Figure 8 (leather), Figure 9 (carpet) and Figure 10 (tile). These results show that for texture, the local model is not sparse: Most Gaussians are used at most positions.

4. Additional heat maps

Additional heat maps for the different classes of the MVTec dataset are shown in Figure 11 and Figure 12.

5. Robustness

We present the detailed results on the randomized MVTec dataset for our method in Table 1, for MahaAD [3] in Table 2, for PaDim [1] in Table 3, for CFlow-ad [2] in Table 4 and for DRÆM [4] in Table 5.

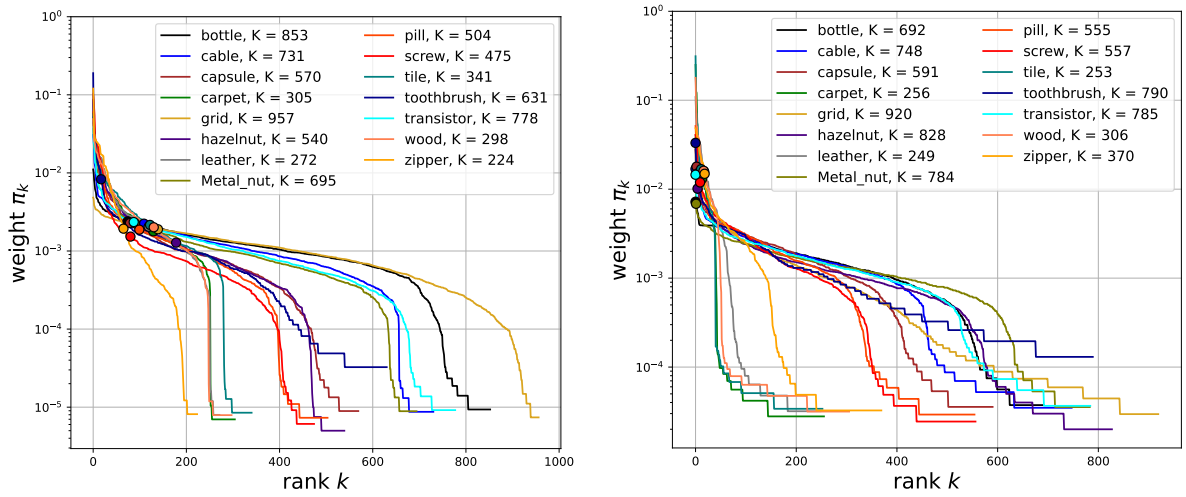

Figure 1. Decay of the global weights π_k in the global mixtures for the second (left, $d = 512$) and third (right, $d = 1024$) layers of WideResNet50-2. The dot indicates the Gaussian from which $\pi_k < (d+1)/|\mathcal{C}_k|$i. From that point the Gaussians' covariance matrices are degenerate. The exact number of Gaussians kept for each class is indicated in the top right of the plot.

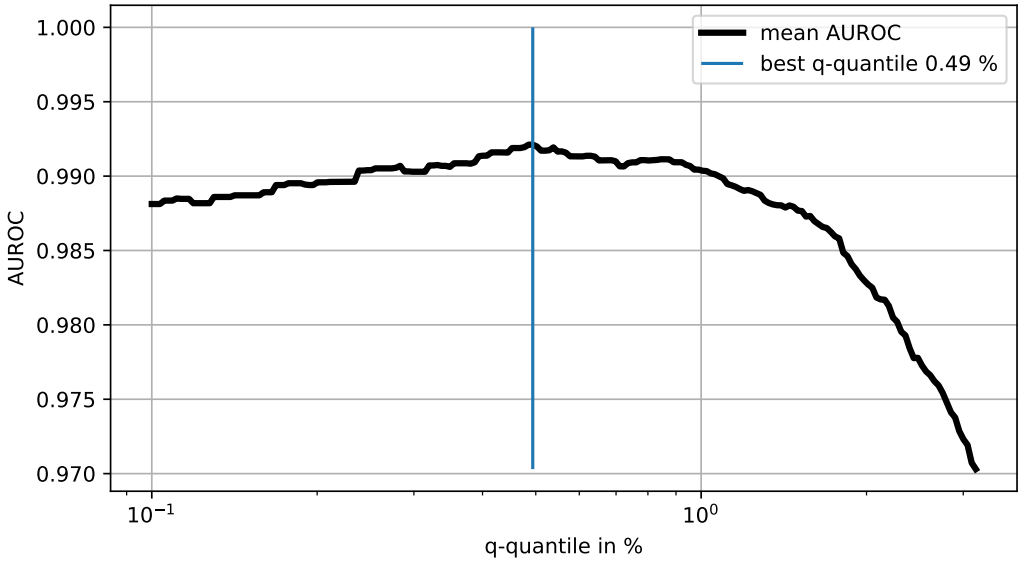

Figure 2. Average Auroc with WideResnet50-2 features, on MvTec AD objects and texture, with respect to the q-quantile chosen for scoring and detection at the image level. The highest AUROC is for a quantile of 0.49%.

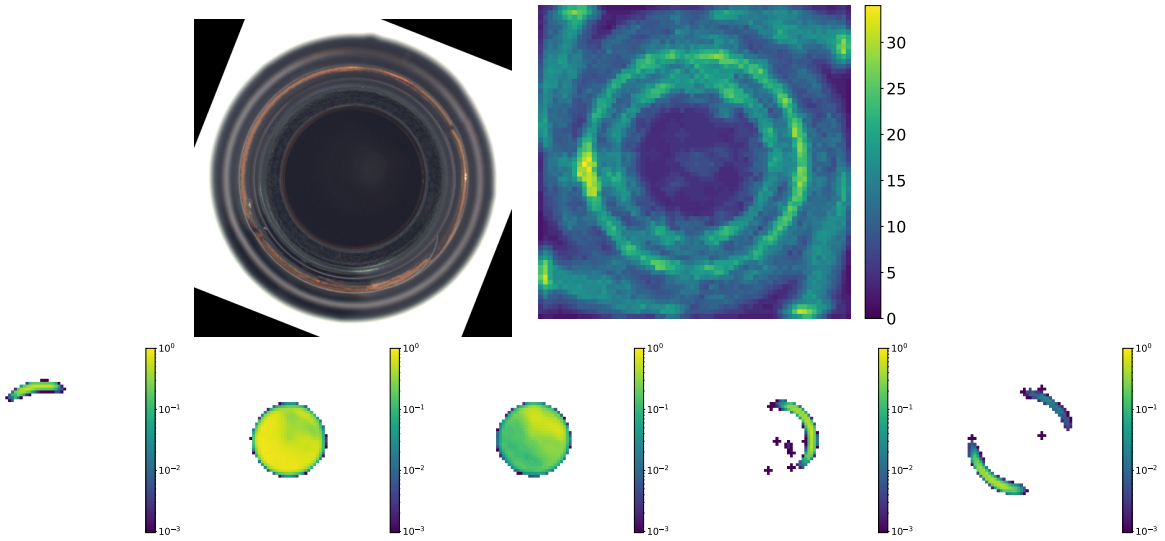

Figure 3. Example of a local weight map for bottle with WideResnet50-2 layer 1. Top row: the image of the bottle (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

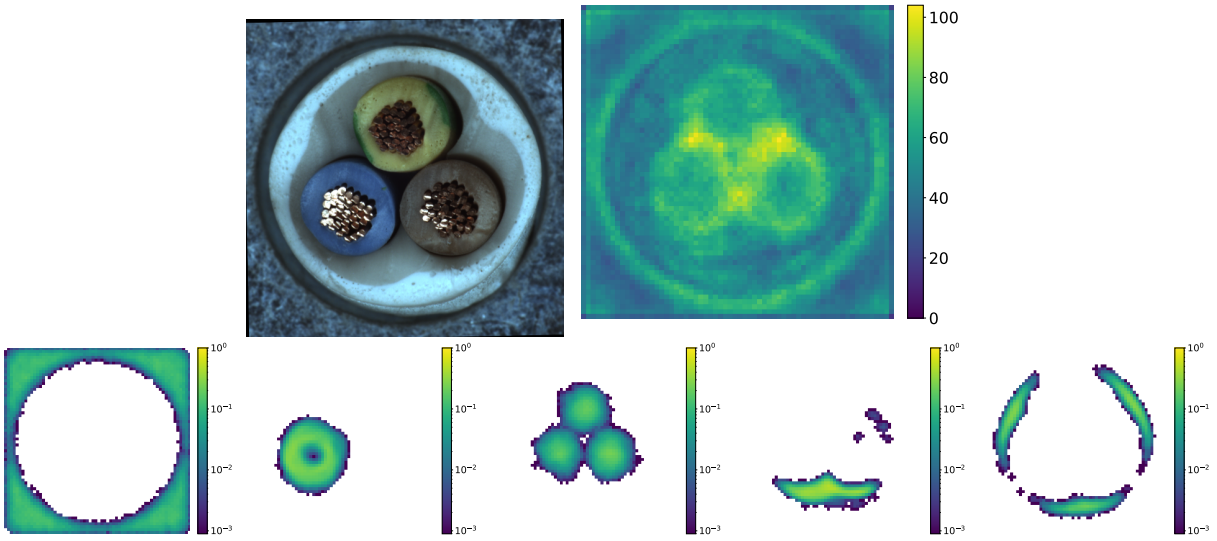

Figure 4. Example of a local weight map for cable with WideResnet50-2 layer 1. Top row: the image of the cable (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

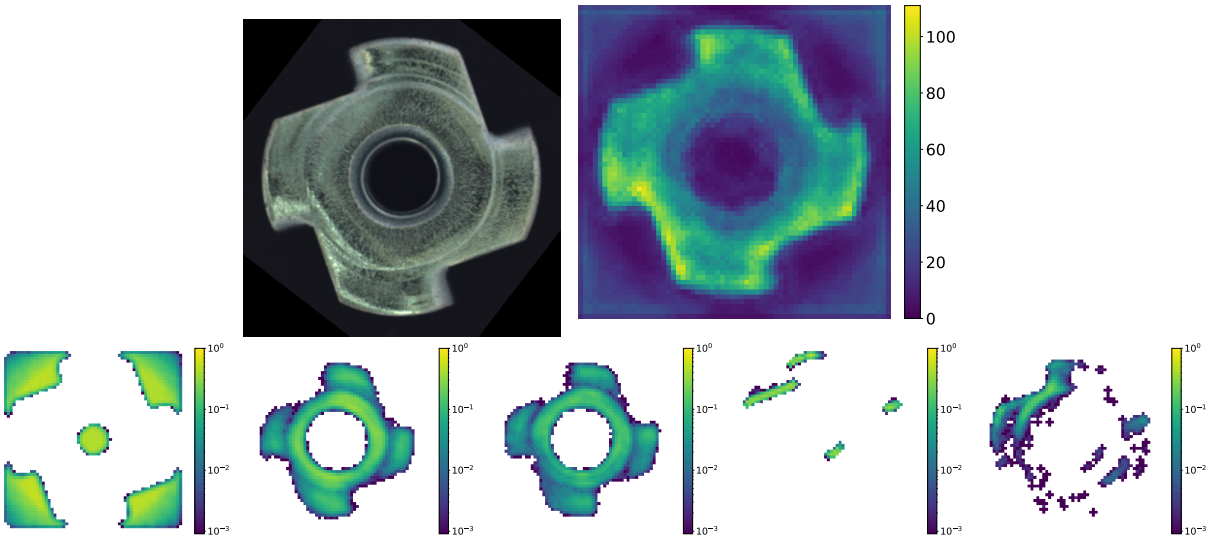

Figure 5. Example of a local weight map for metal nut with WideResnet50-2 layer 1. Top row: the image of the metal nut (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

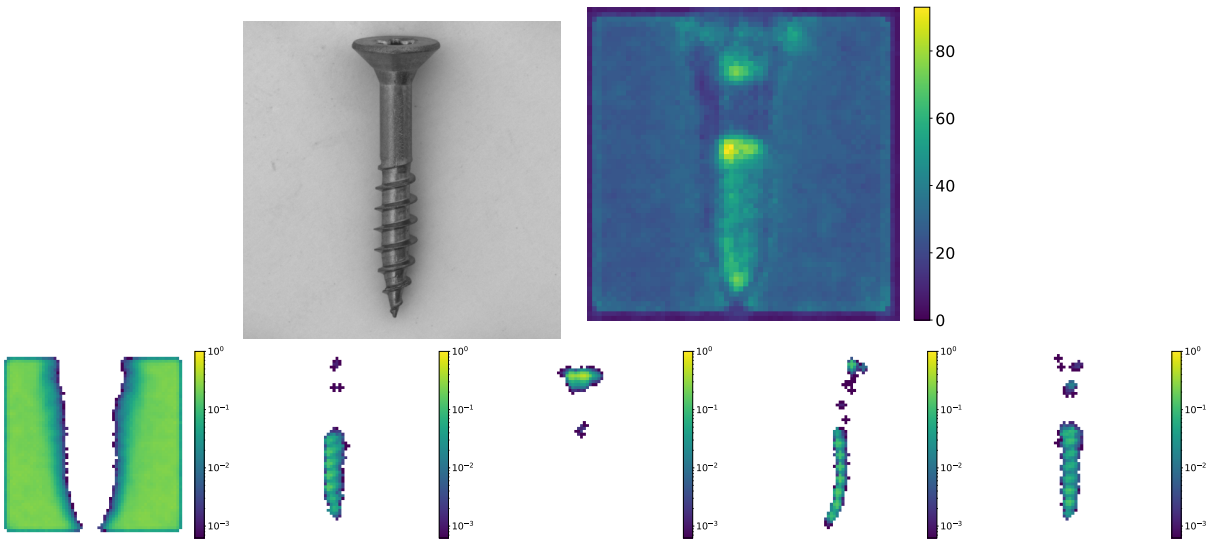

Figure 6. Example of a local weight map for screw with WideResnet50-2 layer 1. Top row: the image of the screw (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

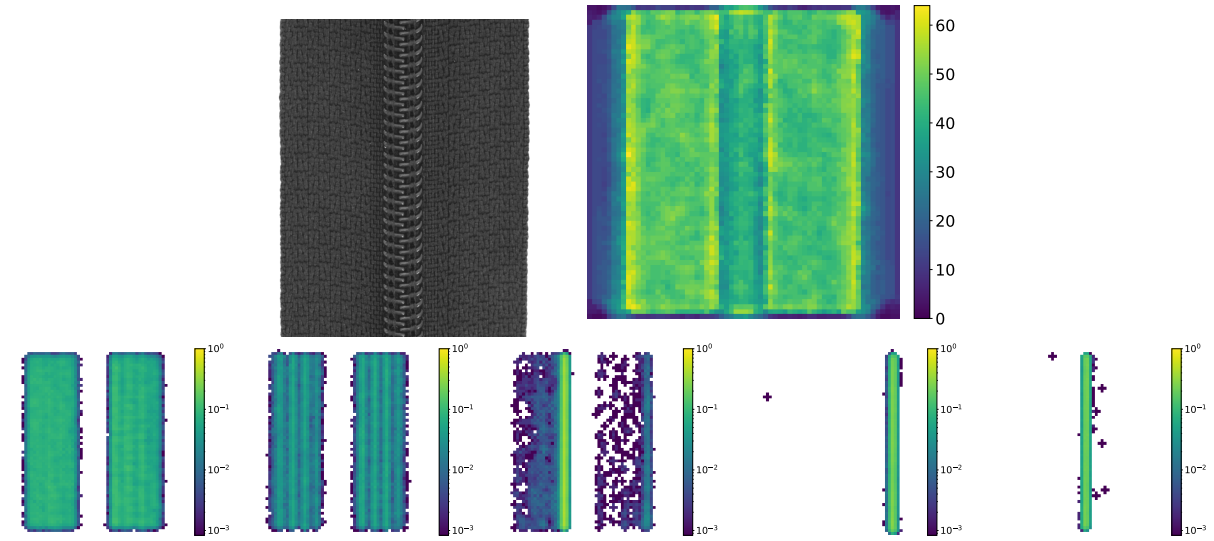

Figure 7. Example of a local weight map for zipper with WideResnet50-2 layer 1. Top row: the image of the zipper (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

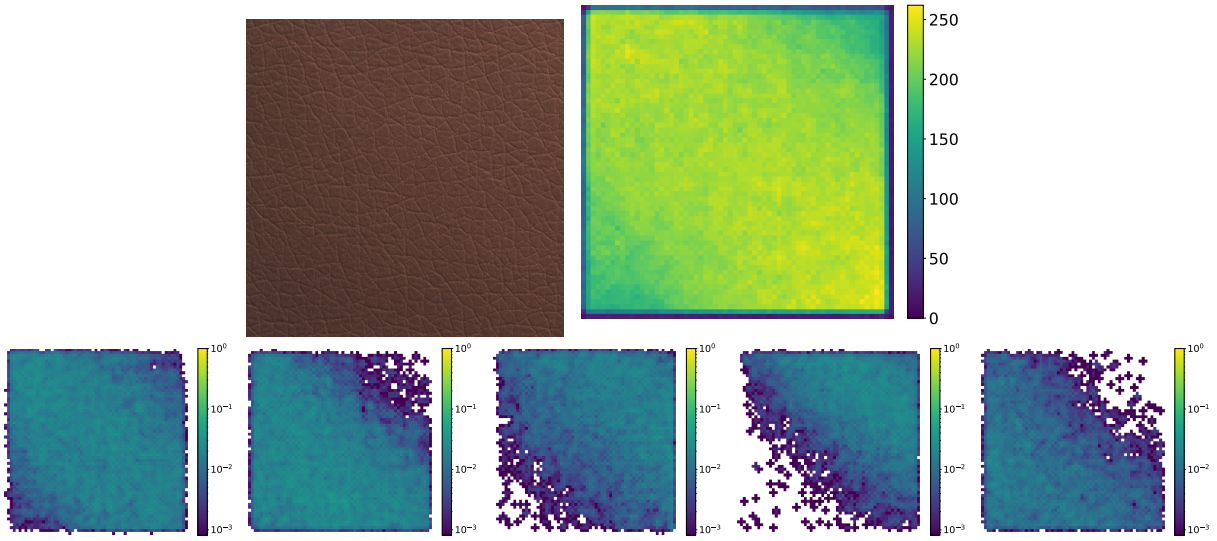

Figure 8. Example of a local weight map for leather with WideResnet50-2 layer 1. Top row: the image of leather (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

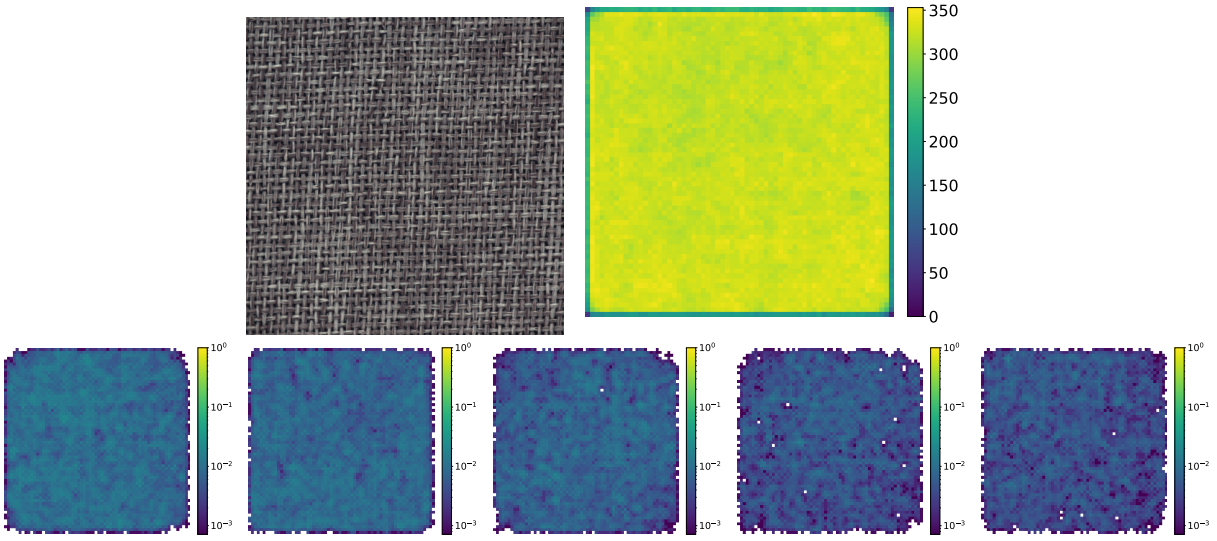

Figure 9. Example of a local weight map for carpet with WideResnet50-2 layer 1. Top row: the image of carpet (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

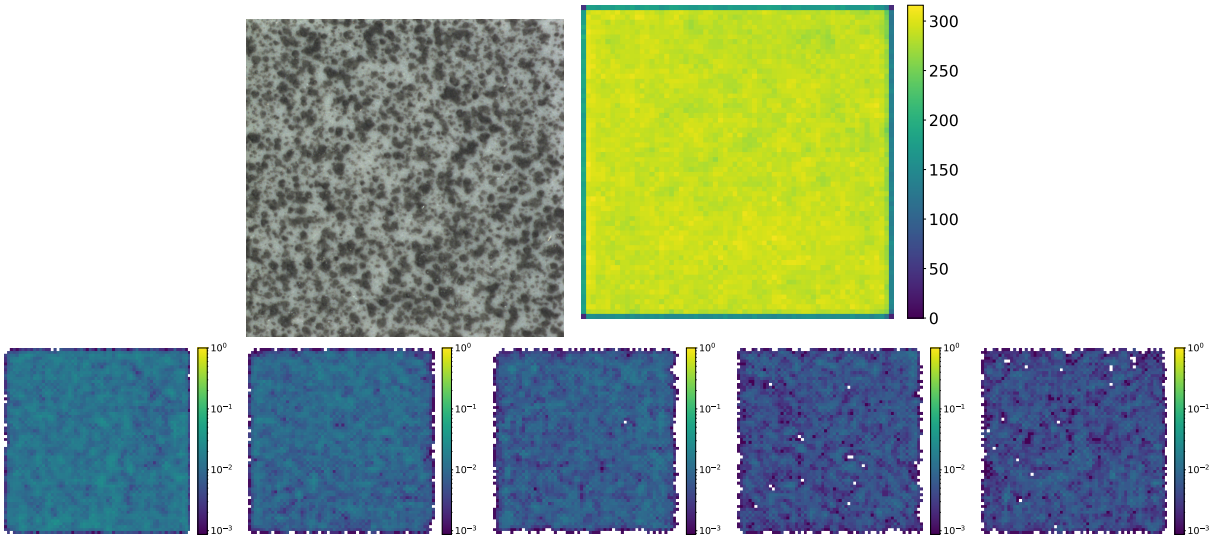

Figure 10. Example of a local weight map for tile with WideResnet50-2 layer 1. Top row: the image of tile (left) and the sparsity map corresponding to the number of non-zero weights per position (right). Bottom: five examples of local weight maps corresponding each to a single Gaussian.

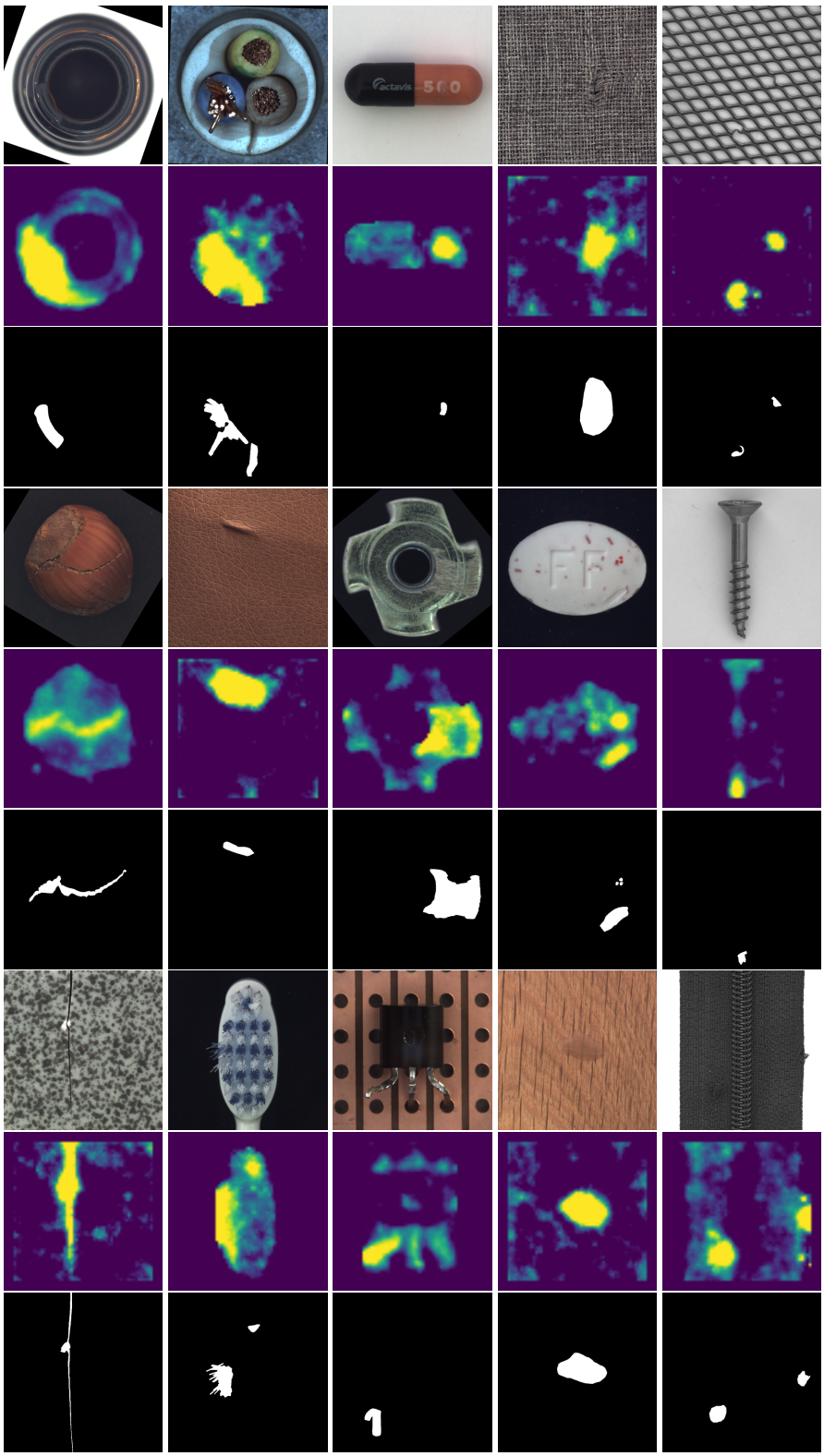

Figure 11. Other examples of heat maps obtained with the multi-scale probabilities on WideResnet50-2.

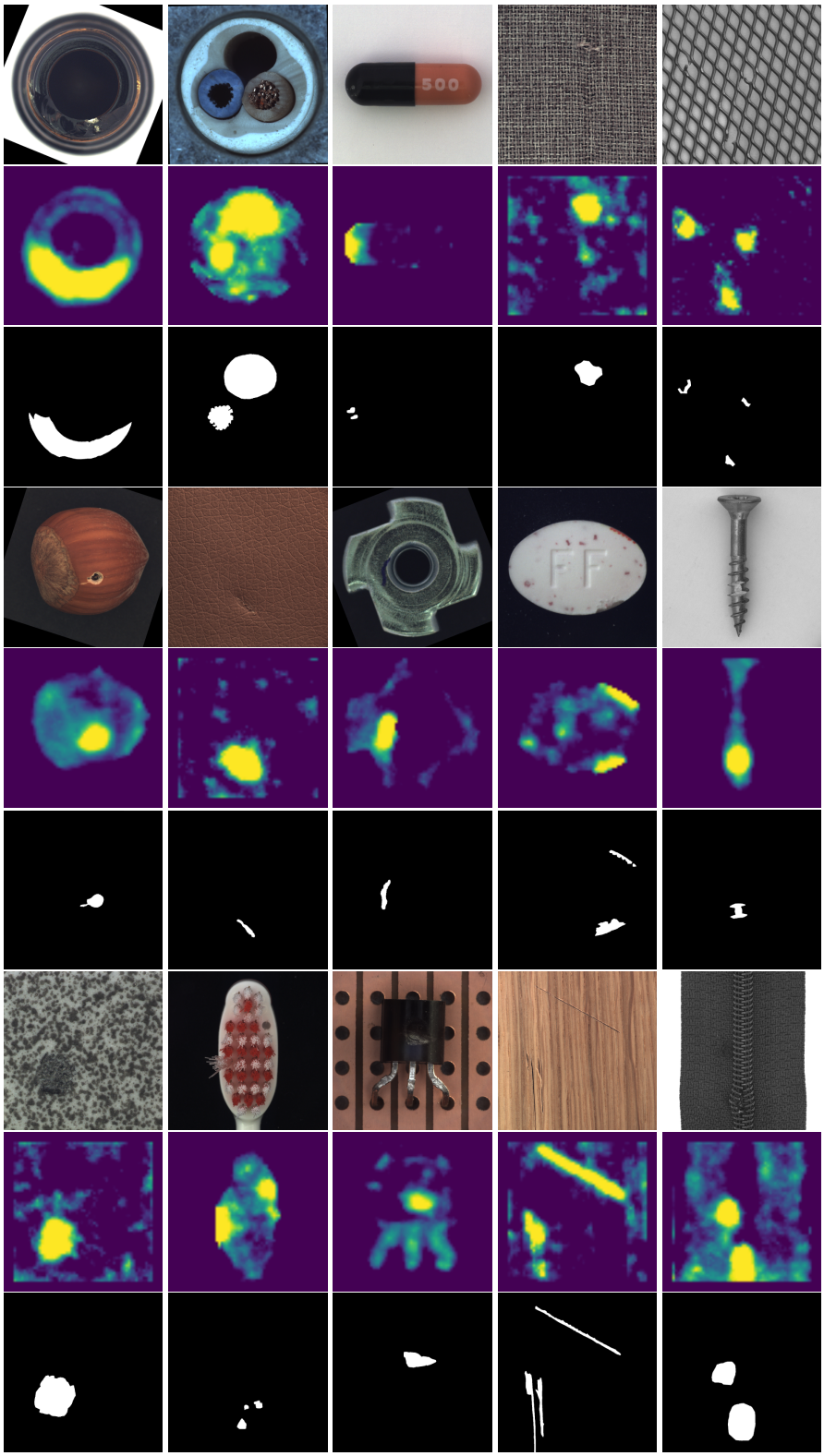

Figure 12. Other examples of heat maps obtained with the multi-scale probabilities on WideResnet50-2.

	MVTech	Rd-MVTech (average)	Rd-MVTech 1	Rd-MVTech 2	Rd-MVTech 3	Rd-MVTech 4	Rd-MVTech 5
bottle	100	100	100	100	100	99.8	100
cable	99.8	92.2	92.2	92.4	92.1	91.6	92.5
capsule	97.8	96.8	94.3	95.6	98.9	97.1	98.0
hazelnut	99.8	99.7	99.9	98.9	99.9	99.9	99.9
metal nut	99.3	99.1	99.1	99.2	99.0	99.2	99.0
pill	96.3	96.6	95.6	97.3	95.5	97.2	97.4
screw	97.9	96.9	96.2	97.9	96.6	96.8	96.8
toothbrush	100	100	100	100	100	100	100
Average	98.9	97.7	97.2	97.7	97.8	97.7	98.0

Table 1. Detailed AUROC for GLAD on five randomized versions of MVTech.

	MVTech	Rd-MVTech (average)	Rd-MVTech 1	Rd-MVTech 2	Rd-MVTech 3	Rd-MVTech 4	Rd-MVTech 5
bottle	100	99.9	100	99.9	99.7	100	100
cable	95.0	81.9	77.7	84.2	81.4	81.7	84.7
capsule	92.5	51.1	50.7	60.7	48.4	45.0	50.8
hazelnut	98.8	98.1	98.4	98.8	96.2	98.4	98.9
metal nut	95.0	89.4	91.6	87.4	89.1	91.1	87.6
pill	84.7	59.5	61.2	63.4	57.5	57.1	58.2
screw	79.8	53.3	50.6	55.0	52.9	46.1	61.7
toothbrush	96.4	57.5	64.2	61.4	38.6	65.3	57.8
Average	92.8	73.8	74.3	76.3	70.5	73.1	75.0

Table 2. Detailed AUROC for MahaAD [3] on five randomized versions of MVTech.

	MVTech	Rd-MVTech (average)	Rd-MVTech 1	Rd-MVTech 2	Rd-MVTech 3	Rd-MVTech 4	Rd-MVTech 5
bottle	99.8	99.6	99.7	99.5	99.4	99.8	99.7
cable	92.2	75.9	77.1	74.8	75.0	78.3	74.4
capsule	91.5	74.9	77.9	74.0	75.3	77.2	70.3
hazelnut	93.3	95.8	95.6	97.0	94.2	95.3	96.8
metal nut	99.2	95.3	96.0	96.5	93.6	96.4	94.0
pill	94.4	79.4	82.2	79.3	81.0	74.2	80.2
screw	84.4	64.4	64.7	67.8	69.2	55.5	65.0
toothbrush	97.2	75.9	76.1	79.2	60.8	81.1	82.2
Average	94.0	82.7	83.7	83.5	81.1	82.2	82.8

Table 3. Detailed AUROC for PaDim [1] on five randomized versions of MVTech.

	MVTech	Rd-MVTech (average)	Rd-MVTech 1	Rd-MVTech 2	Rd-MVTech 3	Rd-MVTech 4	Rd-MVTech 5
bottle	99.8	100	99.9	100	100	100	100
cable	97.1	90.6	93.3	92.7	86.6	89.2	91.1
capsule	98.6	89.6	89.6	88.3	89.2	89.4	91.4
hazelnut	99.3	99.9	100	99.8	99.8	99.9	100
metal nut	99.7	98.6	98.7	98.6	98.2	99.0	98.7
pill	99.1	88.5	91.1	85.6	90.8	86.5	88.7
screw	99.6	81.7	83.1	87.0	80.3	78.3	79.9
toothbrush	99.1	97.6	98.3	98.9	94.2	98.3	98.3
Average	99.0	93.3	94.3	93.9	92.4	92.6	93.5

Table 4. Detailed AUROC for CFlow-ad [2] on five randomized versions of MVTech.

	MVTech	Rd-MVTech (average)	Rd-MVTech 1	Rd-MVTech 2	Rd-MVTech 3	Rd-MVTech 4	Rd-MVTech 5
bottle	99.2	98.2	98.3	99.0	97.9	97.5	98.4
cable	91.8	84.5	84.1	84.0	83.8	84.6	86.2
capsule	98.5	91.5	92.9	91.3	92.2	89.8	91.2
hazelnut	100	100	100	100	100	100	100
metal nut	98.7	98.4	98.6	98.0	97.4	98.9	99.3
pill	98.9	95.3	95.7	94.6	96.5	95.0	94.9
screw	93.9	94.2	96.3	93.3	94.1	92.2	95.0
toothbrush	100	99.9	100	100	100	99.7	100
Average	97.6	95.2	95.7	95.0	95.2	94.7	95.6

Table 5. Detailed AUROC for DRÆM [4] on five randomized versions of MVTech.

	Bottle	*Cable*	*Capsule*	*Hazelnut*	*Metal nut*	*Pill*	*Screw*	*Toothbrush*	*Transistor*	*Zipper*	*Carpet*	*Grid*	*Leather*	*Tile*	*Wood*	*Mean*
Without alignment	100	99.7	97.8	99.9	99.9	96.3	91.4	100	99.6	99.9	99	98,7	100	99.6	98.9	98.7
With alignment	100	99.8	-	99.8	99.4	-	97.9	-	-	-	-	-	-	-	-	99.1

Table 6. Impact of the alignment on the AUROC on the different categories of MVTech. A dashed entry means that the images of the dataset are already aligned and therefore do not require an additional alignment.

6. Additional ablation studies

We present additional ablation studies in this section. In particular, we measure the impact of the alignment step on the MVTech dataset in Table 6. We also compare the performance of the model with different number of Gaussians in Figure 13. In most cases, more Gaussians means better performance. This figure also shows that only 50 to 60% of the Gaussians are kept at the end of the training when starting with many Gaussians meaning that the model tries to avoid redundancy between Gaussians and therefore produces an efficient model. Finally, we also compare the discrimination power of both EM and K-MLE in Figure 14. It shows that even though K-MLE is more computationally efficient than EM, it discriminate defects and good data as well as EM.

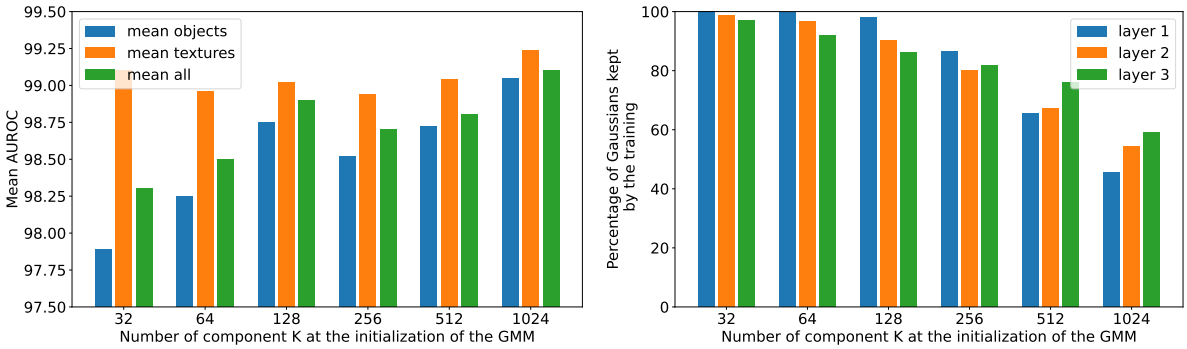

Figure 13. Ablation study on the initial number of components K of the mixture model.

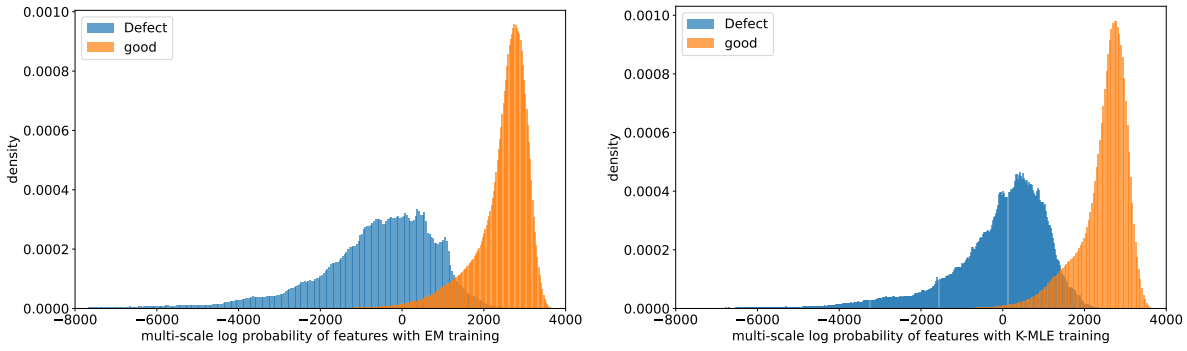

Figure 14. Distribution of multi-scale log likelihood (sum of log likelihoods of each scale) of pixel features on the Mvtec capsule. The left histogram is obtained with EM training while the right is with K-MLE.

References

[1] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: A patch distribution modeling framework for anomaly detection and localization. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges*, pages 475–489, Cham, 2021. Springer International Publishing.

[2] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022.

[3] O. Rippel, P. Mertens, and D. Merhof. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6726–6733, Los Alamitos, CA, USA, jan 2021. IEEE Computer Society.

[4] 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.