Multiresolution Knowledge Distillation for Anomaly Detection
Supplementary Material

Mohammadreza Salehi, Niousha Sadjadi*, Soroosh Baselizadeh*, Mohammad H. Rohban, Hamid R. Rabiee
Department of Computer Engineering, Sharif University of Technology
(smrsalehi, nsadjadi, baselizadeh)@ce.sharif.edu, (rohban, rabiee)@sharif.edu

We report the running time and results of the ablation studies in the paper on different classes of MVTecAD, CIFAR-10, and MNIST in more detail here.

1. Running Time

To show the applicability of our method in real-time applications, we conducted a comprehensive running time test. To avoid any region-based approach in our method, we need only 10 ms for detection on GTX1050 GPU, which is 10X faster than SOTA transformation-based approaches [1, 2]. Besides, as described in the paper, our cloner network is much simpler than VGG-16. Therefore, the overall memory usage at test time is approximately 1.2 times smaller than the VGG-16 size, and hence roughly 20% less than the competing methods. Moreover, our method is general and does not restrict the architecture choice. Lighter models can be used when less memory is available.

2. Intermediate Knowledge

Our framework's performance using different layers as critical points for distillation was discussed in Sec. 3.3.1. Here, we provide the class-detailed performance on MVTecAD and MNIST in Table 1 and Table 2. As discussed in the paper, the performance is enhanced when more intermediate hints are considered. Note that the "only the last layer" setting performs roughly the same as a random detector (AUROC=50%) on some MVTecAD classes.

3. Distillation Effect (Compact Cloner)

In this section, we provide the details of the results in Sec. 3.3.2 of the paper. As mentioned in the paper, a more compact cloner network outperforms a cloner network with equal size to the source. In Tables. 3 and 4, we present a class-detailed comparison for MVTecAD and CIFAR-10 datasets.

* Denotes equal contribution.

4. $\mathcal{L}_{dir}$ and $\mathcal{L}_{val}$

In this part, we present a class-detailed report for the effect of each loss component as discussed in Sec. 3.3.3 in the paper. We report the AUROC for all the classes in MVTecAD and CIFAR-10 datasets in Table 3 and Table 4, respectively. As investigated in the paper, $\mathcal{L}_{total}$, which is a combination of the directional and MSE losses, achieves the best performance when both subtle (MVTecAD) and major (CIFAR-10) anomalies are considered. These results highlight the positive impact of considering a direction-wise notion of activations' knowledge in addition to an MSE approach.

5. Localization using Interpretability Methods

Here, we report detailed results of Sec. 3.3.4 in the paper. In Table 5, the AUROC for all MVTecAD classes is shown with and without applying the Gaussian filter. As discussed in the paper, SmoothGrad highlights the anomalous parts better than others, and GBP performs weaker than the others. Anyway, after applying the noise-removing filters, the methods perform almost the same.

References

[1] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. In *Advances in Neural Information Processing Systems*, pages 9758–9769, 2018.

[2] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*, 2020.

Table 1: Class-detailed AUROC of our proposed method using various layers for distillation. More intermediate layers lead to a performance boost in anomaly detection on MVTecAD.

	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	**Mean**
TheLast	99.6	82.6	79.4	72.8	48.9	91.0	83.6	76.1	66.2	59.4	82.6	85.5	87.6	83.9	89.1	79.22
TheLast2	99.2	89.19	76.8	74.1	58.2	96.3	86.6	78.1	75.3	72.2	86.5	83.4	85.2	95.4	89.8	83.02
TheLast4	99.4	98.4	80.5	73.6	95.1	82.7	94.3	79.3	91.6	78.0	89.2	85.6	92.2	83.3	93.2	**87.74**

Table 2: Class-detailed AUROC of our proposed method using various layers for distillation. More intermediate layers lead to a performance boost in anomaly detection on MNIST.

	0	1	2	3	4	5	6	7	8	9	**Mean**
TheLast	97.65	98.87	93.27	95.10	95.19	94.95	97.63	93.14	94.62	92.93	95.33
TheLast2	99.39	99.60	96.80	97.68	97.94	97.10	98.85	96.86	96.6	96.62	97.74
TheLast4	99.82	99.82	97.79	98.75	98.4	98.16	99.43	98.38	98.41	98.1	**98.71**

Table 3: The detailed AUROC of our method using different loss functions and equal/smaller cloner architectures compared to the source. Both reported on MVTecAD classes. A smaller cloner network performs better compared to the cloner with a size equal to that of the source in general. Also, $\mathcal{L}_{total}$ performs well on both cases, while individual directional or Euclidean losses fail in one.

	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	**Mean**
EqualNet	99.2	88.0	77.7	80.2	75.6	97.4	93.4	76.3	82.6	65.8	89.4	88.9	84.9	93.6	90.8	85.58
SmallerNet	99.4	89.2	80.5	79.3	78.0	98.4	95.1	73.6	82.7	83.3	91.6	92.2	85.6	94.3	93.2	**87.74**
Dir Loss	99.4	89.3	78.8	74.1	50.1	98.1	92.3	81.6	77.8	63.5	91.2	92.0	87.7	88.1	92.2	83.74
MSE Loss	99.4	87.6	81.3	81.3	82.3	98.1	94.3	71.6	85.3	92.3	90.7	94.0	83.9	96.0	94.6	88.8
Total Loss	99.4	89.2	80.5	79.3	78.0	98.4	95.0	73.6	82.7	83.3	91.6	92.2	85.6	94.3	93.2	87.74

Table 4: The detailed AUROC of our method using different loss functions and equal/smaller cloner architectures compared to the source. Both reported on CIFAR-10 classes. A smaller cloner network performs better compared to a cloner with a size equal to that of the source in general. Also, $\mathcal{L}_{total}$ performs well on both cases while individual directional or Euclidean losses fail in one.

	Airplane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	**Mean**
Equal Net	90.04	89.89	80.97	77.24	86.88	91.38	87.72	84.48	90.80	89.34	86.87
Smaller Net	90.53	90.35	79.66	77.02	86.71	91.40	88.98	86.78	91.45	88.91	87.18
Dir Loss	90.42	91.07	79.41	76.98	86.69	91.72	89.21	87.69	91.36	90.27	87.48
MSE Loss	77.36	61.96	66.75	58.94	83.21	60.38	81.67	67.17	79.29	64.88	70.16
Total Loss	90.53	90.35	79.66	77.02	86.71	91.40	88.98	86.78	91.45	88.91	87.2

Table 5: Pixel-wise (AUROC) of anomaly localization on MVTecAD using different interpretability methods with and without Gaussian filtering. Without applying the filters, SmoothGrad performs the best. With Gaussian filtering, however, the methods perform almost the same.

	Gradients + Gaussian	Gradients	SmoothGrad + Gaussian	SmoothGrad	GBP + Gaussian	GBP
Bottle	96.32	93.2	96.03	93.91	95.08	90.46
Cable	82.4	76.24	85.64	81.3	80.21	72.34
Capsule	95.86	93.06	95.55	93.45	95.43	91.53
Carpet	95.64	90.97	95.48	92.98	94.95	90.2
Grid	91.78	84.07	91.4	86.44	90.44	81.46
Hazelnut	94.62	91.3	94.33	89.96	95.06	91.09
Leather	98.05	95.41	98.04	96.76	97.96	94.32
Metal nut	86.38	82.15	86.15	82.54	83.45	77.73
Pill	89.63	86.33	88.99	85.07	90.32	84.99
Screw	95.96	93.42	94.34	91.3	95.3	93.03
Tile	82.77	77.4	82.92	79.37	82.6	76.47
Toothbrush	96.12	92.13	95.64	92.14	95.3	90.28
Transistor	76.45	71.02	76.54	73.13	76.49	68.84
Wood	84.8	78.53	83.4	78.95	84.85	77.47
Zipper	93.9	87.23	93.64	87.18	93.81	85.51
Mean	90.71	86.16	90.54	86.97	90.08	84.38