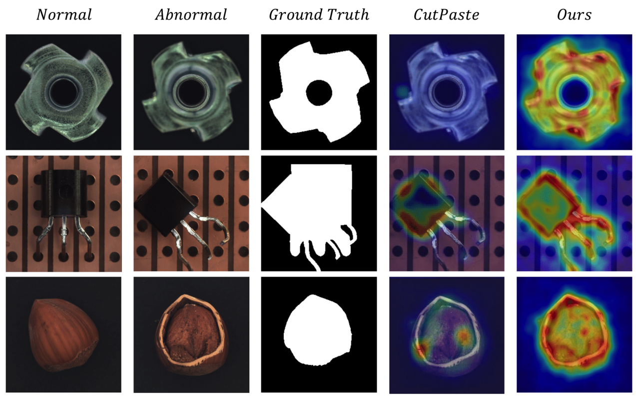

Figure 6. Anomalies from top to bottom: *"flip"* on *"metal nut"*, *"misplaced"* on *"transistor"* and *"crack"* on *"hazelnut"*. Normal samples are provided as reference.

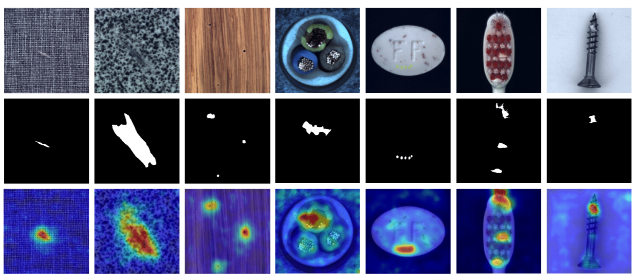

Figure 7. Visualization on tiny or inconspicuous anomalies. From left to right: *carpet*, *tile*, *wood*, *cable*, *pill*, *toothbrush*, and *screw*.

Method	MNIST	F-MNIST	CIFAR10	Caltech-256
LSA [1]	97.5	92.2	64.1	-
OCGAN [27]	97.3	87.8	65.7	-
HRN [17]	97.6	92.8	71.3	-
DAAD [16]	99.0	-	75.3	-
MKD [33]	98.7	94.5	84.5	-
G2D [28]	-	-	-	95.7
OiG [45]	-	-	-	98.2
Ours	**99.3**	**95.0**	**86.5**	**99.9**

Table 4. AUROC(%) results for One-Class Novelty Detection. The best results are marked in bold.

of scores in the similarity map. The baselines in this experiment include LSA [1], OCGAN [27], HRN [17], DAAD [16] and MKD [33]. We also include the comparision with OiG [45] and G2D [28] on **Caltech-256 [13]**.

Tab. 4 summarizes the quantitative results on the three datasets. Remarkably, our approach produces excellent results. Details of the experiments and the results of per-class comparisons are provided in the *Supplementary Material*.

4.3. Ablation analysis

We investigate effective of OCE and MFF blocks on AD and reports the numerical results in Tab. 5. We take the pre-trained residual block [14] as baseline. Embedding from pre-trained residual block may contain anomaly features, which decreases the T-S model's representation discrepancy. Our trainable OCE block condenses feature codes and the MFM block fuses rich features into embedding, allows for more accurate anomaly detection and localization.

Metric	Pre	Pre+OCE	Pre+OCE+MFM
$AUROC_{AD}$	96.0	97.9	**98.5**
$AUROC_{AL}$	96.9	97.4	**97.8**
RPO	91.2	92.4	**93.9**

Table 5. Ablation study on pre-trained bottleneck, OCE, and MFF.

Tab. 6 displays qualitative comparisons of different backbone networks as the teacher model. Intuitively, a deeper and wider network usually have a stronger representative capacity, which facilitates detecting anomalies precisely. Noteworthy that even with a smaller neural network such as ResNet18, our reverse distillation method still achieves excellent performance.

Backbone	ResNet18	ResNet50	WResNet50
$AUROC_{AD}$	97.9	98.4	**98.5**
$AUROC_{AL}$	97.1	97.7	**97.8**
RPO	91.2	93.1	**93.9**

Table 6. Quantitative comparison with different backbones.

Besides, we also explored the impact of different network layers on anomaly detection and shown the results in Tab. 7. For single-layer features, M^2 yields the best result as it trades off both local texture and global structure information. Multi-scale feature fusion helps to cover more types of anomalies.

Score Map	M^1	M^2	M^3	$M^{2,3}$	$M^{1,2,3}$
$AUROC_{AD}$	90.1	97.5	97.2	98.0	98.5
$AUROC_{AL}$	94.0	96.9	96.9	97.6	97.8
RPO	88.6	92.6	89.5	93.2	93.9

Table 7. Ablation study on multi-scale feature distillation.

5. Conclusion

We proposed a novel knowledge distillation paradigm, reverse distillation, for anomaly detection. It holistically addressed the problem in previous KD-based AD methods and boosted the T-S model's response on anomalies. In addition, we introduced trainable one-class embedding and multi-scale feature fusion blocks in reverse distillation to improve one-class knowledge transfer. Experiments showed that our method significantly outperformed previous arts in anomaly detection, anomaly localization, and novelty detection.