G^{2}SF: Geometry-Guided Score Fusion for Multimodal Industrial Anomaly Detection

Supplementary Material

A. Anomaly Synthesis

The whole procedure of our cutPaste anomaly synthesis is illustrated in Fig. 1. More generated anomalous RGB images and associated 3D point clouds are provided in Fig. 7. The synthetic anomalies are pronounced than real-world anomalies in MVTec-3D AD.

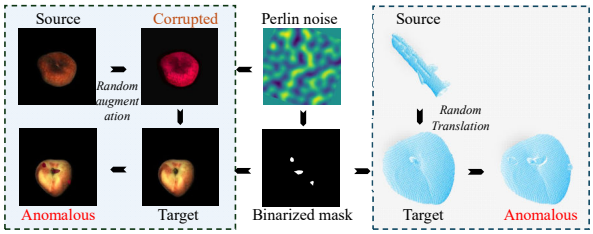

Figure 1. Flowchart of the anomaly synthesis procedure based on CutPaste: (1) Berlin noise maps are binarized into masks to select paste locations; (2) Source regions from normal 3D point clouds/RGB images (or generic models) are cut; (3) Random corruptions (pixel/geometry augmentations) are applied; (4) Modified regions are pasted onto target data.

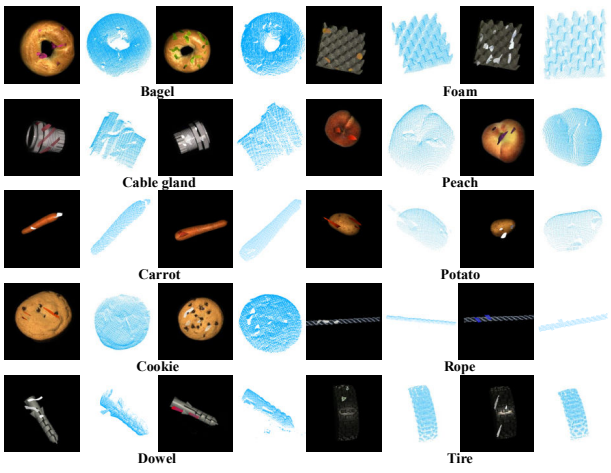

Figure 2. Visualization of generated anomalous RGB images and 3D point clouds on MVTec-3D AD.

B. Implementation Details

B.1. Feature Extraction

RGB features are extracted using DINO ViT-B/8 [7] pretrained on ImageNet [13], generating $28 \times 28 \times 768$ feature maps upsampled to $56 \times 56 \times 768$. Point clouds are processed via Point-MAE [24] pretrained on ShapeNet [8], grouping 3D points into 1024 clusters (feature dimension 1152) and interpolated to $56 \times 56 \times 1152$ for spatial alignment with image feature maps. We use PatchCore [27] to construct modality-specific memory banks with 10% coreset sampling. Notably, we consider the original $28 \times 28 \times 768$ RGB feature maps instead of the upsampled ones when constructing the memory bank of RGB modality to reduce the memory storage. The above procedure remains the same as done in M3DM [32].

Dataset	α	β	γ	μ	Batch Size
MVTec 3D-AD	10	60	8	20	8192
Eyecandies	10	100	2	0.5	15360

Table 1. Turning parameters and batch size in our experiment.

B.2. Data Processing

To accelerate training and underscore discriminative foreground features, we pre-collect features from 800 synthetic anomalous samples per class (Section 3.7) (1000 for Eyecandies), forming a feature dataset equally split for training and validation. The features originate from foreground regions identified via RANSAC plane fitting on 3D point clouds, where areas exceeding 0.005 distance from the fitted plane are retained following [11, 32]. During evaluation on MVTec 3D-AD, background regions remain included. As background features are excluded during training, we set their scaling factors to unity ($w^m_{i,0} = 1$) in Eq. (3), serving as a linear combination of the original unimodal results. For Eyecandies, background regions are neglected following CFM [11].

B.3. Parameter Setting

For $\mathcal{L}_{cns}$ in Eq. (6), we set $k = 5$ and $\eta_0 = 1.2$. We train our LSPN using an Adam optimizer [21] with learning rate $1.5e^{-4}$ and weight decay $1.5e^{-4}$, combined with l_1 regularization of weight $1e^{-4}$ and drop out with ratio 0.5 over 80 epochs. Global scaling factors σ^m are separately optimized at a higher learning rate of $5e^{-3}$ to accelerate convergence. We initialize LSPN by the default strategy in Pytorch, that is, Kaiming Initialization [18]. We observe $w^m_{i,j} \approx 1$ under this strategy, meeting the requirement in Section 3.4. For modality m, we normalize all Euclidean distances $\{s^m_{i,j}\}$ by their mean over the training dataset. Then, we initialize $\sigma^m = 0.5$ with equal importance from each modality. The tuning parameters about loss components and batch size for

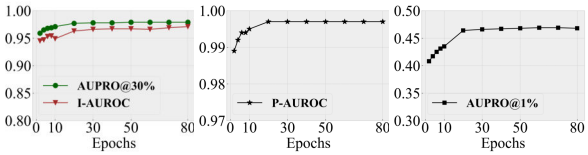

Figure 3. Performance over different training epochs on MVTec 3D-AD dataset.

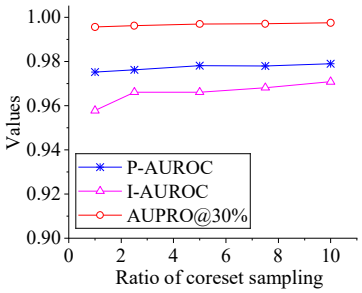

Figure 4. Performance over different ratios of coreset sampling on MVTec 3D-AD dataset.

MVTec 3D-AD and Eyecandies datasets are listed in Table 1. Finally, we implement our G^{2}SF by Pytorch and run all experiments on Linux with a Nvidia RTX 4090 GPU.

C. Learning Curve

We train multiple models that stop at {2,4,6,8,10,20,...,80} epochs respectively, and report their performances (I-AUROC, P-AUROC, AUPRO@30%, and AUPRO@1%) on MVTec-3D AD in Fig. 3. After only 4 epochs, G^{2}SF outperforms M3DM [32] (94.5% I-AUROC, 99.2%P-AUROC, 96.4% AUPRO@30%, and 39.4%AUPRO@1%). This benefits from our design, as our model evolves from an established Euclidean metric to the target anisotropic metric, see Section 3.4.

D. Sensitivity Study on Sizes of Memory Banks

Since the learned metric $l(\cdot)$ is defined in terms of memory prototypes, this subsection presents a sensitivity study of the anomaly detection performance with respect to the number of prototypes. We vary the coreset sampling ratio (see Section B.1) at {10,7.5,5,2.5,1}% levels. At the default 10%, the average numbers of point cloud and RGB image memory prototypes across all categories on the MVTec-3D AD dataset are 83,292 and 20,823, respectively. The corresponding results for the evaluation metrics are summarized in Fig. 4. At the 2.5 % level, G^{2}SF achieves 96.6% I-AUROC, 99.6% P-AUROC, and 97.6% AUPRO@30%. Nevertheless, Fig. 4 demonstrates that the performance of G^{2}SF is relatively stable to the sizes of memory banks.

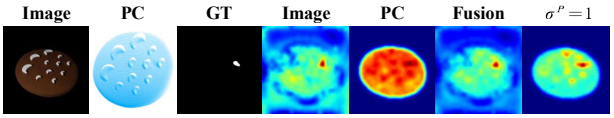

Figure 5. Visualization of learned σ^m and fixed $\sigma^P = 1$ for an example from *ChocolateCookie* category in Eyecandies dataset.

E. Analysis on Global Scaling Factors

G^{2}SF defines trainable global scaling factors σ^m ($m \in P, R$) in Eq. (3) to formulate the metric $l(\cdot)$, where σ^m controls the importance of modality m to final $l(\cdot)$ and normalizes the original Euclidean distances. Fig. 5 demonstrates this mechanism prioritizing image modality over less discriminative point cloud modality of the *ChocolateCookie* category in Eyecandies dataset, compared with fusion and image score maps. Moreover, ignoring the contribution of the RGB image modality by setting $\sigma^P = 1$ and $\sigma^R = 0$ yields significantly degraded results. Furthermore, Table 2 compares performance using learned σ^m against fixed $\sigma^P = \sigma^R = 0.5$ during inference on both the MVTec 3D-AD and Eyecandies datasets. These results collectively underscore the importance of learning optimal σ^m during training.

F. Results on Eyecandies

We provide the quantitative results for all categories in Eyecandies in Table 4. G^{2}SF achieves the best performance of all metrics in terms of +0.5% I-AUROC and +0.8% AUPRO@30% over 2M3DF [1], +0.5% P-AUROC over M3DM [32], +2.2%AUPRO@1%, +2.4%AUPRO@10%, and +2.8%AUPRO@5% over CFM [11]. The qualitative results are illustrated in Fig. 6.

G. Additional Results on MVTec-3D AD

We summarize the results of different score metrics for all categories in MVTec-3D AD in Table 5. The results demonstrate that the final score s_i outperforms unimodal scores $\{s^m_{i,0}\}$ and scaling factors $\{w^m_{i,0}\}$ almost all categories. Additionally, comprehensive quantitative results of various score aggregation operators are provided in 6, which validate the superior performance of $\min$ operator over the majority of classes, owing to its geometric interpretation, as discussed in Section 3.6.

Furthermore, Fig. 7 provides additional qualitative results for all classes in MVTec-3D AD. Consistent with Fig. 5, our G^{2}SF suppresses anomaly scores of normal regions and provides clearer anomaly boundaries, compared to M3DM [32] and CFM [11].

Metric	MVTec-3D AD				Eyecandies			
	I-AUROC	P-AUROC	AUPRO@30%	AUPRO@1%	I-AUROC	P-AUROC	AUPRO@30%	AUPRO@1%
Learn	**0.971**	**0.997**	**0.979**	**0.468**	**0.902**	**0.982**	**0.898**	**0.357**
σ^P=σ^R=0.5	0.964	0.994	0.972	0.460	0.827	0.977	0.892	0.328

Table 2. Ablation study of global scaling factors on MVTec-3D AD and Eyecandies datasets.

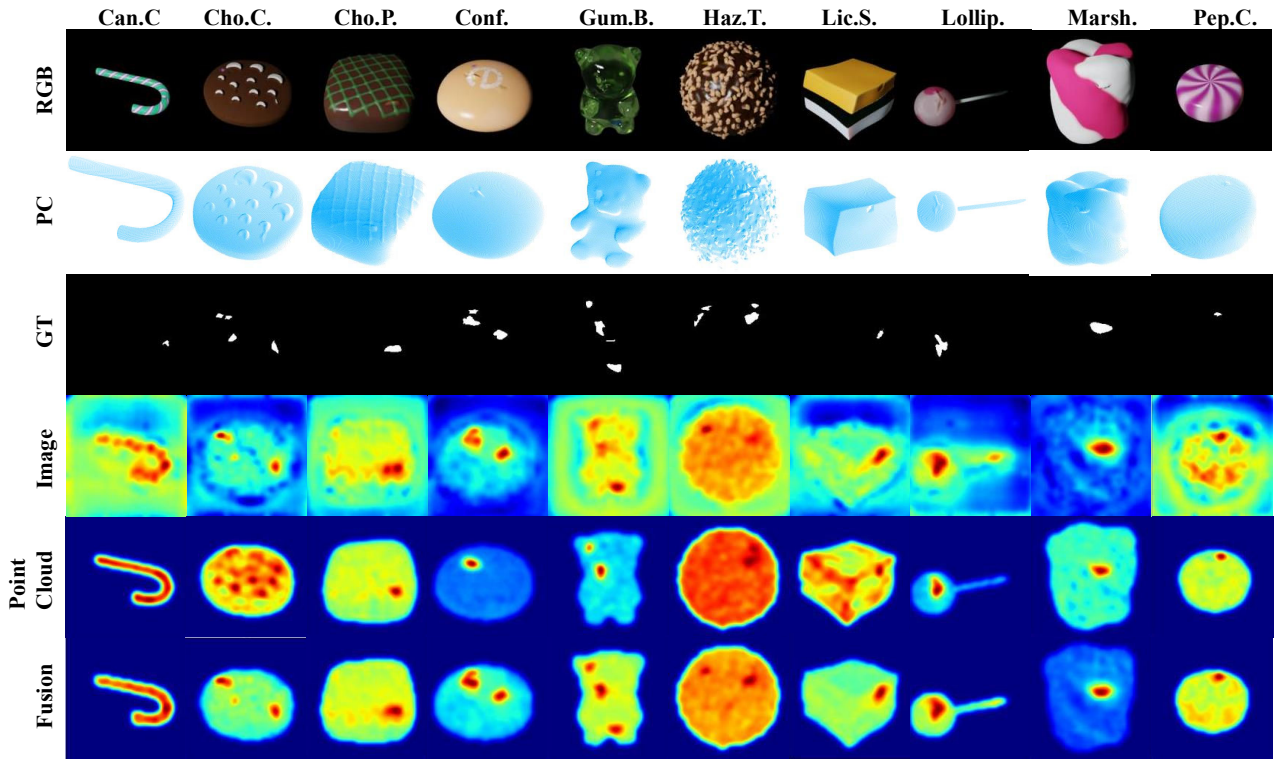

Figure 6. Qualitative results on Eyecandies dataset.

	Method	*Can. C.*	*Cho. C.*	*Cho. P.*	*Conf.*	*Gum. B.*	*Haz. T.*	*Lic. S.*	*Lollip.*	*Marsh.*	*Pep. C.*	**Mean**
I-AUROC	AST[28]	0.574	0.747	0.747	0.889	0.596	0.617	0.816	0.841	0.987	0.987	0.780
	M3DM* [32]	0.597	0.954	0.931	0.990	0.883	0.666	0.923	0.888	0.995	**1.000**	0.882
	CFM [11]	0.680	0.931	0.952	0.880	0.865	0.782	0.917	0.840	0.998	0.962	0.881
	2M3DF [1]	**0.753**	0.955	0.937	0.967	0.901	**0.792**	0.889	**0.913**	0.980	0.893	0.897
	Ours	0.696	**0.963**	**0.967**	**1.000**	**0.907**	0.701	0.939	0.855	0.989	**1.000**	**0.902**
P-AUROC	AST[28]	0.763	0.960	0.911	0.969	0.788	0.837	0.918	0.924	0.983	0.968	0.902
	M3DM* [32]	0.968	0.986	0.964	**0.998**	0.976	0.928	0.976	0.988	0.996	0.995	0.977
	CFM [11]	**0.985**	0.984	0.961	0.986	0.958	0.937	0.968	0.981	0.994	0.978	0.973
	LSFA [30]	0.969	0.957	0.967	0.996	0.971	0.938	0.970	0.990	0.998	**0.987**	0.974
	Ours	0.981	0.982	**0.977**	**0.998**	**0.982**	0.936	**0.981**	0.989	0.997	**0.998**	**0.982**
AUPRO@30%	AST[28]	0.514	0.835	0.714	0.905	0.587	0.590	0.736	0.769	0.918	0.878	0.744
	M3DM* [32]	0.889	0.921	0.808	0.982	0.889	0.675	0.872	0.901	0.964	0.973	0.887
	CFM [11]	**0.942**	0.902	0.831	0.965	0.875	0.762	0.791	0.913	0.939	0.949	0.887
	2M3DF [1]	0.924	**0.935**	0.820	0.940	0.875	**0.781**	0.816	**0.923**	0.958	0.926	0.890
	Ours	0.928	0.897	**0.843**	0.982	**0.890**	0.687	**0.888**	0.913	**0.971**	**0.981**	**0.898**
AUPRO@1%	AST[28]	0.035	0.230	0.129	0.234	0.092	0.069	0.139	0.090	0.255	0.224	0.149
	M3DM* [32]	0.166	0.388	0.329	0.486	0.315	0.131	0.323	0.258	0.462	0.454	0.331
	CFM [11]	**0.229**	0.397	0.345	0.389	0.353	**0.188**	0.333	0.236	0.455	0.428	0.335
	Ours	0.174	**0.416**	**0.377**	**0.487**	**0.360**	0.164	**0.353**	**0.281**	**0.478**	**0.479**	**0.357**
AUPRO@10%	AST[28]	0.285	0.709	0.545	0.770	0.404	0.350	0.584	0.544	0.770	0.744	0.570
	M3DM* [32]	0.677	0.836	0.698	**0.947**	0.754	0.410	0.732	0.712	0.913	0.924	0.760
	CFM [11]	**0.827**	0.815	0.731	0.896	0.741	**0.550**	0.663	**0.739**	0.893	0.868	0.772
	Ours	0.784	**0.820**	**0.752**	0.946	**0.780**	0.490	**0.787**	**0.739**	**0.922**	**0.943**	**0.796**
AUPRO@5%	AST[28]	0.173	0.592	0.421	0.635	0.288	0.242	0.461	0.378	0.634	0.617	0.444
	M3DM* [32]	0.479	**0.759**	0.626	**0.894**	0.655	0.300	0.634	0.562	0.849	0.861	0.661
	CFM [11]	**0.662**	0.750	0.653	0.801	0.657	**0.427**	0.609	0.552	0.838	0.796	0.675
	Ours	0.578	**0.759**	**0.687**	0.892	**0.703**	0.399	**0.698**	**0.568**	**0.858**	**0.889**	**0.703**

Table 3. Results on Eyecandies. Best results in **bold**, runner-ups underlined.

	Method	*Bagel*	*Cable Gland*	*Carrot*	*Cookie*	*Dowel*	*Foam*	*Peach*	*Potato*	*Rope*	*Tire*	**Mean**
AUPRO@1%	BTF[19]	0.428	0.365	0.452	0.431	0.370	0.244	0.427	0.470	0.298	0.345	0.383
	AST[28]	0.388	0.322	0.470	0.411	0.328	0.275	0.474	0.487	0.360	0.474	0.398
	Shape-guided[10]	-	-	-	-	-	-	-	-	-	-	0.456
	M3DM [32]	0.414	0.395	0.447	0.318	**0.422**	0.335	0.444	0.351	0.416	0.398	0.394
	CFM [11]	0.459	0.431	**0.485**	0.469	0.394	0.413	0.468	0.487	0.464	**0.476**	0.455
	Ours	**0.481**	**0.443**	0.484	**0.471**	0.410	**0.468**	**0.487**	**0.499**	**0.468**	0.471	**0.468**
.@10%	CFM [11]	0.937	0.917	0.947	0.897	0.855	0.906	0.942	0.947	0.926	**0.944**	0.922
	Ours	**0.946**	**0.928**	**0.947**	**0.937**	**0.912**	**0.932**	**0.946**	**0.950**	**0.933**	0.942	**0.937**
.@5%	CFM [11]	0.877	0.843	**0.894**	0.840	0.765	0.828	0.884	0.894	0.865	**0.889**	0.858
	Ours	**0.892**	**0.860**	0.893	**0.877**	**0.828**	**0.873**	**0.892**	**0.899**	**0.870**	0.885	**0.877**

Table 4. Additional **AUPRO@1%**, **AUPRO@10%** and **AUPRO@5%** results on MVTec-3D AD. Best results in **bold**, runner-ups underlined.

	Method	*Bagel*	*Cable Gland*	*Carrot*	*Cookie*	*Dowel*	*Foam*	*Peach*	*Potato*	*Rope*	*Tire*	**Mean**
I-AUROC	$s^R_{i,0}$	0.945	**0.939**	0.914	0.733	0.939	0.768	0.944	0.621	0.930	0.755	0.849
	$s^P_{i,0}$	0.951	0.637	0.981	0.952	0.837	0.786	0.923	0.930	0.858	0.688	0.854
	$w^R_{i,0}$	0.969	0.732	0.989	0.928	0.831	0.983	0.981	**0.988**	0.965	0.798	0.916
	$w^P_{i,0}$	0.969	0.727	0.989	0.930	0.829	0.983	0.982	**0.988**	0.963	0.796	0.915
	s_i	**0.997**	0.923	**0.993**	**0.967**	**0.966**	**0.991**	**0.994**	**0.988**	**0.966**	**0.922**	**0.971**
P-AUROC	$s^R_{i,0}$	0.992	0.993	0.995	0.976	0.996	0.955	0.994	0.991	0.995	0.995	0.988
	$s^P_{i,0}$	0.987	0.945	0.997	0.940	0.981	0.935	0.993	0.996	0.994	0.981	0.975
	$w^R_{i,0}$	0.964	0.910	0.995	0.774	0.954	0.973	0.965	0.998	0.997	0.952	0.948
	$w^P_{i,0}$	0.993	0.983	0.998	0.916	0.986	0.991	0.995	**0.999**	0.998	0.977	0.984
	s_i	**0.998**	**0.995**	**0.999**	**0.996**	**0.996**	**0.997**	**0.998**	**0.999**	**0.999**	**0.998**	**0.997**
AUPRO@30%	$s^R_{i,0}$	0.954	0.973	0.974	0.887	**0.974**	0.847	0.971	0.957	0.965	0.970	0.947
	$s^P_{i,0}$	0.96	0.806	0.977	0.899	0.929	0.761	0.971	0.976	0.946	0.927	0.915
	$w^R_{i,0}$	0.957	0.900	0.976	0.724	0.896	0.911	0.944	0.982	0.974	0.898	0.916
	$w^P_{i,0}$	0.979	0.953	0.981	0.908	0.934	0.950	0.978	**0.983**	0.976	0.939	0.958
	s_i	**0.982**	**0.976**	**0.982**	**0.979**	0.971	**0.976**	**0.982**	**0.983**	**0.978**	**0.981**	**0.979**
AUPRO@1%	$s^P_{i,0}$	0.367	0.423	0.416	0.235	**0.415**	0.296	0.399	0.285	0.396	0.399	0.363
	$s^P_{i,0}$	0.424	0.063	0.448	0.406	0.243	0.238	0.385	0.409	0.356	0.222	0.319
	$w^R_{i,0}$	0.480	0.384	0.480	0.378	0.356	0.469	0.477	0.494	0.455	0.416	0.439
	$w^P_{i,0}$	**0.481**	0.378	0.481	0.421	0.356	**0.474**	0.480	0.494	0.453	0.415	0.443
	s_i	**0.481**	**0.443**	**0.484**	**0.471**	0.410	0.468	**0.487**	**0.499**	**0.468**	**0.471**	**0.468**

Table 5. Results of various score metrics on MVTec-3D AD. Best results in **bold**, runner-ups underlined.

	Method	*Bagel*	*Cable Gland*	*Carrot*	*Cookie*	*Dowel*	*Foam*	*Peach*	*Potato*	*Rope*	*Tire*	**Mean**
I-AUROC	$l_{i,0}$	0.997	0.925	0.989	0.962	0.953	**0.993**	0.996	0.964	0.960	0.892	0.963
	max	**0.998**	0.932	0.982	0.966	0.949	0.968	**0.999**	0.862	0.953	0.922	0.953
	mean	0.997	**0.940**	0.988	**0.975**	0.954	**0.993**	0.995	0.980	0.960	**0.941**	**0.972**
	min	0.997	0.923	**0.993**	0.967	**0.966**	0.991	0.994	**0.988**	**0.966**	0.922	0.971
P-AUROC	$l_{i,0}$	0.997	0.993	**0.999**	0.994	0.995	**0.997**	**0.998**	**0.999**	**0.999**	0.997	**0.997**
	max	0.980	0.989	**0.999**	0.917	0.993	0.986	**0.998**	**0.999**	0.998	0.996	0.986
	mean	0.977	0.992	**0.999**	0.888	0.995	0.983	0.997	**0.999**	**0.999**	**0.998**	0.983
	min	**0.998**	**0.995**	**0.999**	**0.996**	**0.996**	**0.997**	**0.998**	**0.999**	**0.999**	**0.998**	**0.997**
AUPRO@30%	$l_{i,0}$	0.981	0.970	**0.982**	0.974	0.967	0.975	**0.982**	**0.983**	**0.978**	0.979	0.977
	max	0.968	0.960	**0.982**	0.883	0.954	0.936	**0.982**	**0.983**	0.976	0.974	0.960
	mean	0.965	0.970	**0.982**	0.815	0.961	0.932	**0.982**	**0.983**	**0.978**	0.979	0.955
	min	**0.982**	**0.976**	**0.982**	**0.979**	**0.971**	**0.976**	**0.982**	**0.983**	**0.978**	**0.981**	**0.979**
AUPRO@1%	$l_{i,0}$	0.480	0.429	0.482	0.444	**0.411**	0.467	0.485	0.497	0.465	0.448	0.461
	max	**0.481**	0.390	**0.485**	0.428	0.391	0.467	**0.494**	0.495	0.457	0.443	0.453
	mean	0.480	0.436	**0.485**	0.396	0.403	**0.469**	0.491	0.497	0.465	0.460	0.458
	min	**0.481**	**0.443**	0.484	**0.471**	0.410	0.468	0.487	**0.499**	**0.468**	**0.471**	**0.468**

Table 6. Results of various score aggregation strategies on MVTec-3D AD. Best results in **bold**, runner-ups underlined.

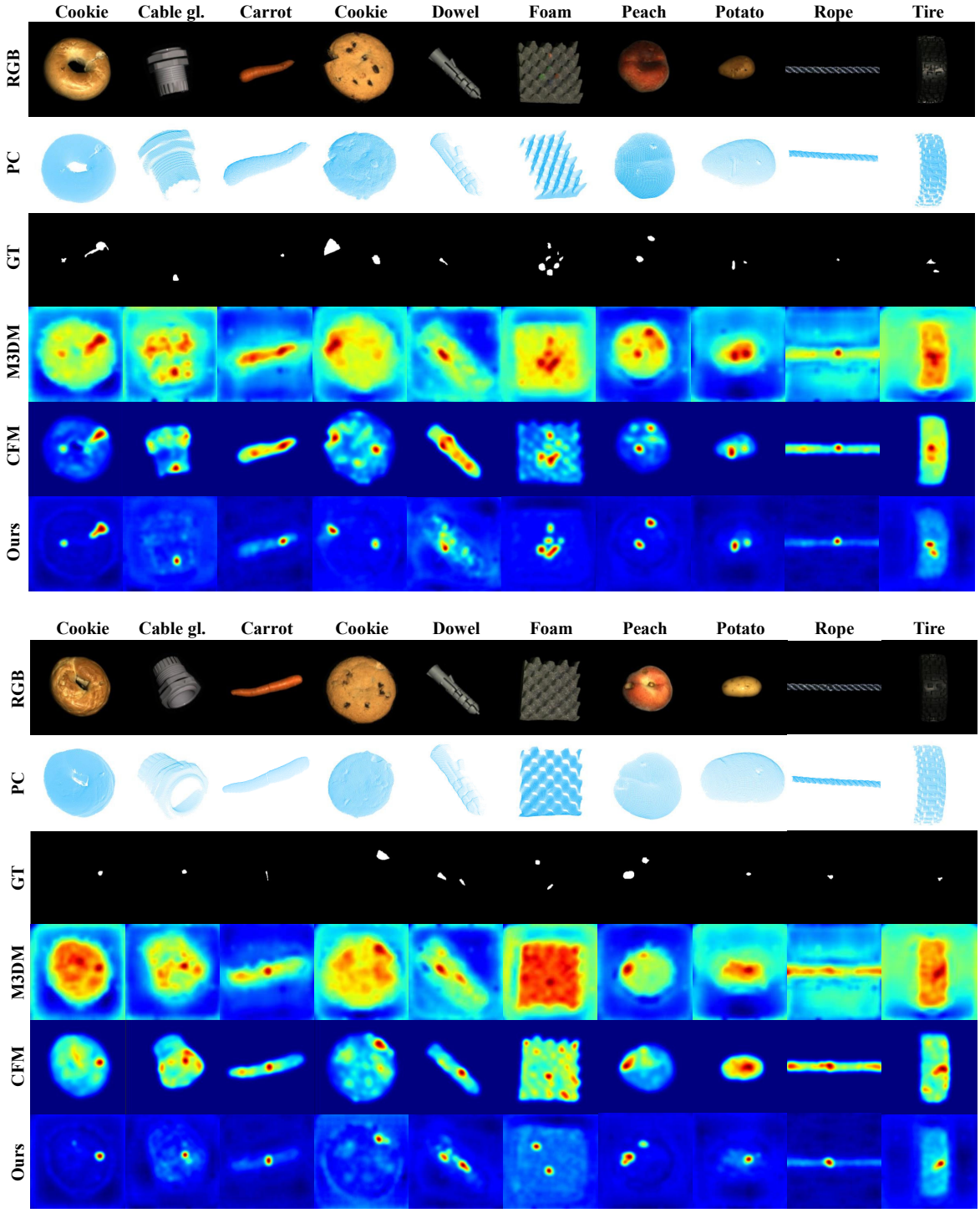

Figure 7. Additional qualitative results on MVTec-3D AD dataset.