Multimodal Industrial Anomaly Detection via Hybrid Fusion (Supplementary Material)

Yue Wang[1]*, Jinlong Peng[2]*, Jiangning Zhang[2], Ran Yi[1]†, Yabiao Wang[2], Chengjie Wang[2,1]
[1]Shanghai Jiao Tong University, Shanghai, China; [2]Youtu Lab, Tencent
[1]{imwangyue,ranyi}@sjtu.edu.cn [2]{jeromepeng,vtzhang,caseywang,jasoncjwang}@tencent.com

Overview

This supplementary material includes:

- The implementation details about the hardware and software packages (Sec. A);
- P-AUROC score for anomaly segmentation (Sec. B);
- Detailed score for ablation study of all categories of MVTec-3D AD (Sec. C);
- Detailed score for Point Transformer setting of all categories of MVTec-3D AD (Sec. D);
- Detailed score for few-shot setting of all categories of MVTec-3D AD (Sec. E);
- Discussion about Backbone Choices (Sec. F);
- The results of all categories of Eyecandies (Sec. G);
- The visualization results of all categories of MVTec-3D AD (Sec. H).

A. Implementation Details

We implement M3DM with Pytorch[1] and Scikit-Learn package[2]. The feature extractors and memory banks algorithm are based on Pytorch and we use the Scikit-Learn package for OCSVM [10]. The AUROC calculation also relies on Scikit-Learn package. All experiments are run on a single Nvidia Tesla V100 and cost at most 50 GB of memories for the full setting.

*Equal contributions. This work was done when Yue Wang was a intern at Tencent Youtu Lab.
†Corresponding author.
[1]https://pytorch.org/
[2]https://scikit-learn.org/

B. P-AUROC for Segmentation

In the main paper, we report the AUPRO score for anomaly segmentation. In this section, we report the P-AUROC score to further verify the segmentation performance of our method, as shown in Tab. I. We mainly compare our results with FPFH [6], PatchCore [8] and AST [9][3]. For the multimodal input, we get the same score as PatchCore + FPFH method and is 1.6% higher than the AST. For single RGB input, we still have a 2% improvement over PatchCore. For 3D segmentation, similar to the AUPRO results reported in the main paper, our 3D segmentation results are a little bit lower than the FPFH-based method, and we believe this is also caused by the bias between the label and the point clouds we discuss in Section 4.7 in the main paper. The P-AUROC is a saturated metric for anomaly segmentation, and the difference between methods is smaller than the difference in AUPRO.

C. Detailed Results of Ablation Study

In the main paper Section 4.3, we conduct ablation studies on UFF, DLF, and multiple memory banks. In this section, we report the detailed ablation study results of all categories of MVTec-3D AD. Tab. II and Tab. III separately illustrate the I-AUPRO and AUPRO scores with the following settings: 1) Only Point Clouds ($\mathcal{M}_{pt}$) information; 2) Only RGB ($\mathcal{M}_{rgb}$) information; 3) Single memory bank ($\mathcal{M}_{fs}$) directly concatenating Point Transformer feature and RGB feature together; 4) Single memory bank ($\mathcal{M}_{fs}$) using UFF to fuse multimodal features; 5) Building two memory banks ($\mathcal{M}_{rgb}, \mathcal{M}_{pt}$) separately and directly adding the scores together; 6) Building two memory banks separately ($\mathcal{M}_{rgb}, \mathcal{M}_{pt}$) and using DFL for the final result; 7) Building three memory banks ($\mathcal{M}_{rgb}, \mathcal{M}_{pt}, \mathcal{M}_{fs}$) (Ours). With the UFF, the Foam, Cookie, and Peach have a great improvement to the single domain input and the w/o UFF version, which means the UFF encourages the interaction between multimodal features and creates useful in-

[3]Since AST [9] only provided the mean score in its paper, we simply illustrate the mean score of AST.

	Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
3D	FPFH [6]	0.994	0.966	0.999	0.946	0.966	0.927	0.996	0.999	0.996	0.990	**0.978**
	Ours	0.981	0.949	0.997	0.932	0.959	0.925	0.989	0.995	0.994	0.981	0.970
RGB	PatchCore [8]	0.983	0.984	0.980	0.974	0.972	0.849	0.976	0.983	0.987	0.977	0.967
	Ours	0.992	0.990	0.994	0.977	0.983	0.955	0.994	0.990	0.995	0.994	**0.987**
RGB+3D	AST [9]	-	-	-	-	-	-	-	-	-	-	0.976
	PatchCore + FPFH [6]	0.996	0.992	0.997	0.994	0.981	0.974	0.996	0.998	0.994	0.995	**0.992**
	Ours	0.995	0.993	0.997	0.985	0.985	0.984	0.996	0.994	0.997	0.996	**0.992**

Table I. P-AUROC score for anomaly segmentation of all categories of MVTec-3D AD [1] dataset. The P-AUROC is a saturated metric for anomaly segmentation, and the difference between methods is smaller than the AUPRO.

formation for anomaly detection and segmentation. With double memory banks, the Carrot, Cookie and Potato score have an improvement, and the DLF help improve the hard categories such as Cable Gland and Tire. With three memory banks, most advantages of DLF and UFF have been maintained, and our full setting gets the best results, which indicates DLF and UFF complements each other and jointly achieves the best performance.

D. Detailed Results of PFA Analysis

In the main paper Section 4.4, we conduct experiments on PFA settings. Here we report the detailed results of the main paper Table 5 with scores on each category in Tab. IV and Tab. V. The PFA settings are:

- Two important hyper-parameters of Point Transformer: the number of groups and the groups' size during farthest point sampling; The number of groups decides how many features will be extracted by the Point Transformer and the groups' size is equal to the concept of the receptive field;
- 3D anomaly detection experiment with the original point groups feature: a point group can be seen as a *patch*, and the memory bank store point groups feature here; The detection method is as same as the patch-based one, and to get the segmentation predictions, we first project point group feature to a 2D plane and use an inverse distance interpolation to get every pixel value.

We found that directly calculating the anomaly on the point groups has some advantage in certain categories (e.g. Bagel, Cable Gland, Foam and Rope), and the reason is that after furthest point sampling (FPS) the original point group feature contains more small defects information. As the patch gets smaller, our 2D plane point feature gets a better performance in detecting the small defects.

E. Detailed Results of Few-shot Setting

In the main paper Section 4.6, we evaluate our method on Few-shot settings, and the detailed results on all of the categories are illustrated in Tab. VII and Tab. VI. We randomly select 10 and 5 images from each category as training data and test the few-shot model on the full testing dataset. We find that our method in a 10-shot or 5-shot setting still has a better segmentation performance than some non-few-shot methods. In the 50-shot setting, We found that some categories get better performance than the full dataset version (e.g. Bagel and Potato), which means the memory bank building algorithm still has some improvement space, and we will discuss the problem in future research.

F. Backbone Choices

Feature extractors play an important role in anomaly detection. In this section, we explore different backbone settings on both point cloud and RGB images. For RGB we compare four extractor settings: 1) A ViT-B/8 supervised backbone pretrained with ImageNet [5] 1K; 2) A ViT-B/8 supervised backbone pretrained with ImageNet 21K; 3) A ViT-S/8 self-supervised backbone pertrained via DINO [3]; 4) A ViT-B/8 self-supervised backbone pertrained via DINO. And for Point Clouds transformer, we compare two self-supervised pretrained backbones: 1) Point-Bert [12]; 2) Point-MAE [7]. The detection and segmentation results are separately illustrated in Tab. VIII and Tab. IX. The results show that the self-supervised pretrained methods have better results than the supervised ones, and small backbones pretrained with self-supervised methods perform better than the bigger ones pretrained with supervised methods. The performance of Point-MAE is better than that of Point-Bert, we think the reason is that Point-MAE needs to reconstruct more point cloud details than Point-Bert, thus can catch small defects in anomaly detection.

G. Eyecandies Results

We have noticed that recently a new dataset Eyecandies [2] provides multimodel information of 10 categories of candies, and each category contains 1000 images for training and 50 images for public testing. The source dataset provides 6 RGB images, which are in different light conditions, a depth map, and a normal map of each sam-

Method	Memory Banks	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
Only PC	$\mathcal{M}_{pt}$	0.941	0.651	0.965	0.969	0.905	0.760	0.880	0.974	0.926	0.765	0.874
Only RGB	$\mathcal{M}_{rgb}$	0.944	0.918	0.896	0.749	0.959	0.767	0.919	0.648	0.938	0.767	0.850
w/o UFF	$\mathcal{M}_{fs}$	0.920	0.900	0.914	0.727	0.963	0.795	0.946	0.656	0.954	0.792	0.857
w/ UFF	$\mathcal{M}_{fs}$	0.976	0.895	0.922	0.912	0.949	0.868	0.978	0.723	0.960	0.798	**0.898**
w/o DLF	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}$	0.981	0.831	0.980	0.985	0.960	0.905	0.936	0.964	0.967	0.780	0.929
w/ DLF	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}$	0.980	0.880	0.975	0.965	0.947	0.910	0.943	0.927	0.958	0.840	**0.932**
Ours	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}, \mathcal{M}_{fs}$	0.994	0.909	0.972	0.976	0.960	0.942	0.973	0.899	0.972	0.850	**0.945**

Table II. Detailed I-AUROC score for ablation on anomaly detection of all categories of MVTec-3D AD.

Method	Memory Banks	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
Only PC	$\mathcal{M}_{pt}$	0.943	0.818	0.977	0.882	0.881	0.743	0.958	0.974	0.950	0.929	0.906
Only RGB	$\mathcal{M}_{rgb}$	0.952	0.972	0.973	0.891	0.932	0.843	0.970	0.956	0.968	0.966	0.942
w/o UFF	$\mathcal{M}_{fs}$	0.951	0.971	0.974	0.893	0.935	0.855	0.972	0.958	0.969	0.967	0.944
w/ UFF	$\mathcal{M}_{fs}$	0.963	0.964	0.978	0.930	0.946	0.896	0.974	0.966	0.972	0.972	**0.956**
w/o DLF	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}$	0.968	0.925	0.979	0.914	0.909	0.948	0.975	0.976	0.967	0.965	0.953
w/ DLF	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}$	0.965	0.968	0.978	0.933	0.933	0.927	0.976	0.967	0.971	0.973	**0.959**
Ours	$\mathcal{M}_{pt}, \mathcal{M}_{rgb}, \mathcal{M}_{fs}$	0.970	0.971	0.979	0.950	0.941	0.932	0.977	0.971	0.971	0.975	**0.964**

Table III. Detailed AUPRO score for ablation anomaly segmentation of all categories of MVTec-3D AD.

ple. In this section, we convert the Eyecandies dataset to the format supported by M3DM. In detail, we use the environment light image as our input RGB data, and for 3D data, we first convert the depth image to point clouds with internal parameters, then we remove the background points with point coordinates. For computation efficiency, we use only less than 400 samples from each category for training. We try to build memory banks of different sizes (ranging from 10 to 400 samples) to find the best one under this dataset. As illustrated in Tab. X and Tab. XI, we report the best I-AUCROC and P-AUCROC scores. Compared with baseline methods, we have significant improvement in both the RGB setting and RGB+3D setting. Previous work did not report the AUPRO score on the Eyecandies dataset, and for reference in further study, we provide the this segmentation performance metric score of M3DM in Tab. XII.

H. Visualization Results

In this section, we visualize more anomaly segmentation results for all categories of MVTec-3D AD datasets. As shown in Fig. I, we visualize the heatmap results of our method and PatchCore + FPFH, both with multimodal inputs. Compared with PatchCore + FPFH results, our method gets better segmentation maps.

S.G	N.G	Sampling	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
64	784	point groups	0.933	0.579	0.854	0.843	0.874	0.748	0.761	0.863	0.989	0.483	0.793
128	1024	point groups	0.945	0.690	0.905	0.925	0.897	0.809	0.854	0.888	0.991	0.509	0.841
64	784	8×8 patch	0.905	0.508	0.939	0.923	0.817	0.725	0.857	0.916	0.897	0.561	0.805
128	1024	8×8 patch	0.886	0.560	0.925	0.971	0.832	0.711	0.873	0.909	0.897	0.624	0.819
128	1024	4×4 patch	0.941	0.651	0.965	0.969	0.905	0.760	0.880	0.974	0.926	0.765	**0.874**

Table IV. Detailed I-AUROC results of exploring Point Transformer setting on the pure 3D setting. S.G means the point number per group, and N.G means the total number of point groups. We achieve the best performance with 1,024 point groups per sample and each point group contains 128 points; Compared with directly calculating anomaly scores on point groups, the method based on a 2D plane patch needs a small patch size towards high performance.

S.G	N.G	Sampling	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
64	784	point groups	0.906	0.709	0.942	0.854	0.869	0.681	0.871	0.906	0.943	0.447	0.813
128	1024	point groups	0.956	0.812	0.964	0.895	0.892	0.644	0.965	0.974	0.966	0.891	0.896
64	784	8×8 patch	0.899	0.789	0.970	0.848	0.871	0.718	0.931	0.951	0.939	0.874	0.879
128	1024	8×8 patch	0.934	0.808	0.977	0.856	0.877	0.745	0.949	0.970	0.948	0.894	0.896
128	1024	4×4 patch	0.943	0.818	0.977	0.882	0.881	0.743	0.958	0.974	0.950	0.929	**0.906**

Table V. Detailed AUPRO results of exploring Point Transformer setting on the pure 3D setting. S.G means the point number per group, and N.G means the total number of point groups. We get the best performance with 1024 point groups per sample and each point group contains 128 points; Compared with directly calculating segmentation scores on point groups, the method based on a 2D plane patch needs a small patch size towards high performance.

Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
5-shot	0.974	0.645	0.833	0.942	0.636	0.798	0.820	0.781	0.914	0.615	0.796
10-shot	0.987	0.662	0.854	0.969	0.643	0.799	0.908	0.771	0.931	0.682	0.821
50-shot	0.997	0.745	0.957	0.966	0.910	0.915	0.937	0.910	0.946	0.744	0.903
Full dataset	0.994	0.909	0.972	0.976	0.960	0.942	0.973	0.899	0.972	0.850	**0.945**

Table VI. Few-shot I-AUROC of all categories of MVTec-3D AD. Our method still has good anomaly detection performance on few-shot settings.

Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
5-shot	0.959	0.879	0.974	0.906	0.879	0.848	0.968	0.957	0.963	0.935	0.927
10-shot	0.972	0.910	0.976	0.923	0.905	0.870	0.972	0.956	0.967	0.939	0.939
50-shot	0.969	0.955	0.977	0.940	0.906	0.912	0.971	0.965	0.968	0.959	0.952
Full dataset	0.970	0.971	0.979	0.950	0.941	0.932	0.977	0.971	0.971	0.975	**0.964**

Table VII. Few-shot AUPRO of all categories of MVTec-3D AD. Our method on few-shot still has a better anomaly segmentation performance than most non-few-shot methods.

	Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	Supervised ImageNet 1K	0.793	0.729	0.774	0.709	0.723	0.601	0.607	0.606	0.605	0.556	0.670
	Supervised ImageNet 21K	0.814	0.658	0.788	0.630	0.784	0.582	0.615	0.459	0.674	0.621	0.662
	DINO ViT-S/8	0.933	0.865	0.898	0.786	0.878	0.759	0.902	0.520	0.898	0.748	0.819
	DINO ViT-B/8	0.944	0.918	0.896	0.749	0.959	0.767	0.919	0.648	0.938	0.767	**0.850**
3D	Point-Bert	0.900	0.632	0.932	0.915	0.851	0.659	0.826	0.899	0.894	0.530	0.803
	Point-MAE	0.941	0.651	0.965	0.969	0.905	0.760	0.880	0.974	0.926	0.765	**0.874**

Table VIII. I-AUROC score for anomaly detection of MVTec-3D AD [1] dataset with different backbone. For RGB feature extractor, The self-supervised backbone is better than the supervised ones.

	Method	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	Supervised ImageNet 1K	0.844	0.842	0.892	0.681	0.842	0.568	0.765	0.865	0.915	0.871	0.808
	Supervised ImageNet 21K	0.805	0.878	0.927	0.712	0.888	0.62	0.785	0.909	0.919	0.930	0.837
	DINO ViT-S/8	0.948	0.973	0.971	0.906	0.947	0.788	0.972	0.954	0.964	0.949	0.937
	DINO ViT-B/8	0.952	0.972	0.973	0.891	0.932	0.843	0.970	0.956	0.968	0.966	**0.942**
3D	Point-Bert	0.895	0.775	0.972	0.841	0.871	0.680	0.918	0.964	0.938	0.877	0.873
	Point-MAE	0.943	0.818	0.977	0.882	0.881	0.743	0.958	0.974	0.950	0.929	**0.906**

Table IX. AUPRO score for anomaly segmentation of MVTec-3D AD [1] dataset with different backbones. For RGB feature extractor, the self-supervised backbone is better than the supervised ones.

	Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sand-wich	Lollipop	Marshm -allow	Pepper-mint Candy	Mean
3D	Ours	0.482	0.589	0.805	0.845	0.780	0.538	0.766	0.827	0.800	0.822	0.725
RGB	RGB [2]	0.527	0.848	0.772	0.734	0.590	0.508	0.693	0.760	0.851	0.730	0.701
	STFPM [11]	0.551	0.654	0.576	0.784	0.737	**0.790**	0.778	0.620	0.840	0.749	0.708
	PaDiM [4]	0.531	0.816	0.821	0.856	0.826	0.727	0.784	0.665	0.987	0.924	0.794
	Ours	**0.648**	**0.949**	**0.941**	**1.000**	**0.878**	0.632	**0.933**	**0.811**	**0.998**	**1.000**	**0.879**
RGB+3D	RGB-D [2]	0.529	0.861	0.739	0.752	0.594	0.498	0.679	0.651	0.838	0.75	0.689
	RGB-cD-N [2]	0.596	0.843	0.819	0.846	0.833	0.550	0.750	**0.846**	0.940	0.848	0.787
	Ours	**0.624**	**0.958**	**0.958**	**1.000**	**0.886**	**0.758**	**0.949**	0.836	**1.000**	**1.000**	**0.897**

Table X. I-AUROC score for anomaly detection of all categories of Eyecandies [2] dataset. The results of baselines are from the [2].

	Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sand-wich	Lollipop	Marshm -allow	Pepper-mint Candy	Mean
3D	Ours	0.977	0.903	0.902	0.93	0.875	0.832	0.909	0.968	0.868	0.918	0.908
RGB	RGB [2]	**0.972**	0.933	**0.960**	0.945	0.929	0.815	0.855	0.977	0.931	0.928	0.925
	Ours	**0.956**	**0.979**	0.958	**0.998**	**0.976**	**0.941**	**0.977**	**0.986**	**0.997**	**0.988**	**0.976**
RGB+3D	RGB-D [2]	0.973	0.927	0.958	0.945	0.929	0.806	0.827	0.977	0.931	0.928	0.920
	RGB-cD-N [2]	**0.980**	0.979	**0.982**	0.978	0.951	0.853	0.971	0.978	0.985	0.967	0.962
	Ours	0.974	**0.987**	0.962	**0.998**	**0.966**	**0.941**	**0.973**	**0.984**	**0.996**	**0.985**	**0.977**

Table XI. P-AUROC score for anomaly detection of all categories of Eyecandies [2] dataset. The results of baselines are from the [2].

Method	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sand-wich	Lollipop	Marshm -allow	Pepper-mint Candy	Mean
Point Clouds	0.911	0.645	0.581	0.748	0.748	0.484	0.608	0.904	0.646	0.750	0.702
RGB	0.867	0.904	0.805	0.982	0.871	0.662	0.882	0.895	0.970	0.962	0.880
Point Clouds + RGB	0.906	0.923	0.803	0.983	0.855	0.688	0.880	0.906	0.966	0.955	0.882

Table XII. AUPRO score for anomaly detection of all categories of Eyecandies [2] dataset.

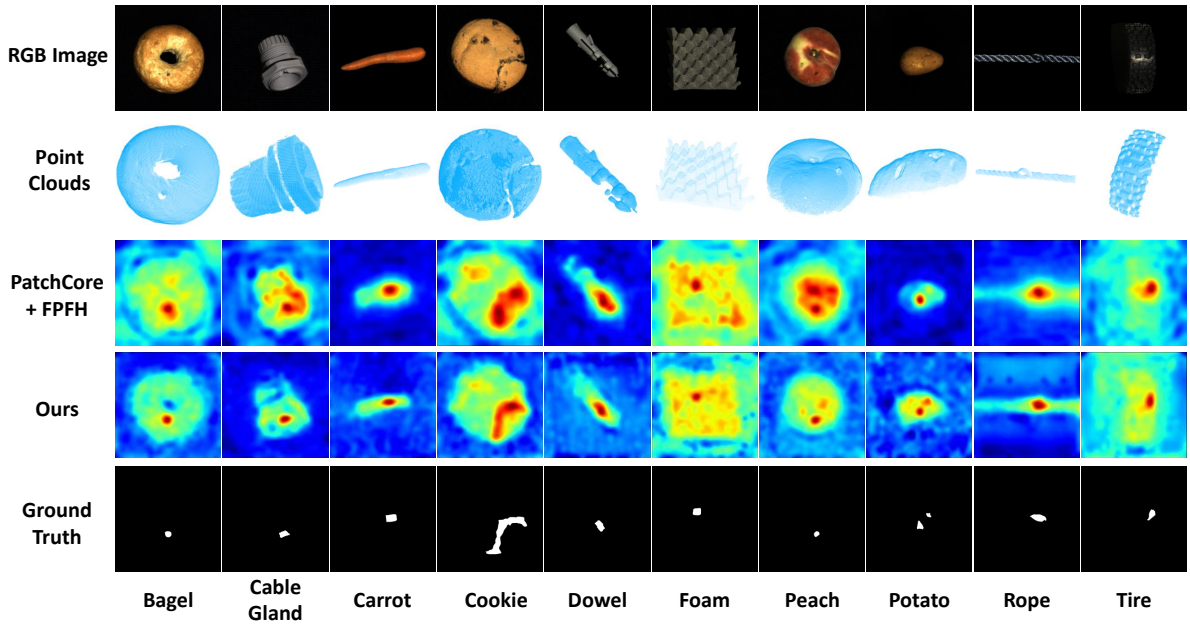

Figure I. Heatmap of our anomaly segmentation results (multimodal inputs). Compared with PatchCore + FPFH, our method outputs a more accurate segmentation region.

References

[1] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In Giovanni Maria Farinella, Petia Radeva, and Kadi Bouatouch, editors, *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 5: VISAPP, Online Streaming, February 6-8, 2022*, pages 202–213. SCITEPRESS, 2022. 2, 4, 5

[2] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the Asian Conference on Computer Vision*, pages 3586–3602, 2022. 2, 5

[3] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9650–9660, 2021. 2

[4] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 5

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[6] Eliahu Horwitz and Yedid Hoshen. An empirical investigation of 3d anomaly detection and segmentation. *arXiv preprint arXiv:2203.05550*, 2022. 1, 2

[7] Yatian Pang, Wenxiao Wang, Francis E. H. Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning, 2022. 2

[8] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 1, 2

[9] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2592–2602, 2023. 1, 2

[10] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001. 1

[11] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. In *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*, page 306. BMVA Press, 2021. 5

[12] Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19313–19322, 2022. 2