

# Student-Teacher Feature Pyramid Matching for Anomaly Detection (Supplementary)

Guodong Wang<sup>\*1,2</sup>  
 wanggd@buaa.edu.cn  
 Shumin Han<sup>\*3</sup>  
 hanshumin@baidu.com  
 Errui Ding<sup>3</sup>  
 dingerrui@baidu.com  
 Di Huang<sup>†1,2</sup>  
 dhuang@buaa.edu.cn

<sup>1</sup> State Key Laboratory of Software Development Environment  
 Beihang University  
 Beijing, China  
<sup>2</sup> School of Computer Science and Engineering  
 Beihang University  
 Beijing, China  
<sup>3</sup> Department of Computer Vision Technology  
 Baidu, Inc.  
 Beijing, China

The supplementary material provides more quantitative and visual results on the MvTec Anomaly Detection (MvTec AD) [1], Shanghai Tech Campus (STC) [2] and CIFAR-10 [3].

## 1 MvTec Anomaly Detection Dataset

For each category (except toothbrush with only one defect type), two different defect types are shown in Figure 3, Figure 4, Figure 5 and Figure 6. It demonstrates that our method enables anomaly detection of any size and any type.

## 2 Shanghai Tech Campus Dataset

SSIM-AE	$l_2$ -AE	CAVGA- $R_u$	SPADE	Ours
0.76	0.74	0.85	0.899	<b>0.913</b>

Table 1: Pixel-level anomaly detection on the STC dataset. The performance is measured by the average AUC-ROC across 12 scenes. The results for all the counterpart approaches are quoted from [4].

We further evaluate our method for pixel-level anomaly detection on the STC dataset. This dataset is originally created for anomaly detection in 12 surveillance scenes. Following [4], we construct the training and test sets by extracting every fifth frame from each

scene. The training set only contains normal images while the test set contains both normal and anomalous images. Different from the MVTec AD dataset, STC focuses more on motion anomalies such as fighting and car intruding captured by surveillance cameras.

Table 1 shows that the performance of our method exceeds that of CAVGA- $R_u$  [10] by a significant margin. Note that CAVGA- $R_u$  reports the state-of-the-art result in the literature on unsupervised anomaly detection. Once again, our method outperforms SPADE in spite of a lightweight network used.

Figure 1 visualizes the results of our method on three anomalous images from the STC dataset. We clearly find that the anomalous regions in the images are very difficult to precisely localize if we only use a specific level of features. Combination of the features in a hierarchical fashion shows a good way to delineate the anomalous regions regardless of their sizes.

### 3 CIFAR-10

OCGAN	1-NN	OC-SVM	$l_2$ -AE	VAE	STAD	Ours
0.657	0.819	0.739	0.790	0.750	0.820	<b>0.832</b>

Table 2: Image-level anomaly detection on CIFAR-10. The performance is measured by the average AUC-ROC across 10 categories. The results for all the counterpart approaches are quoted from [10].

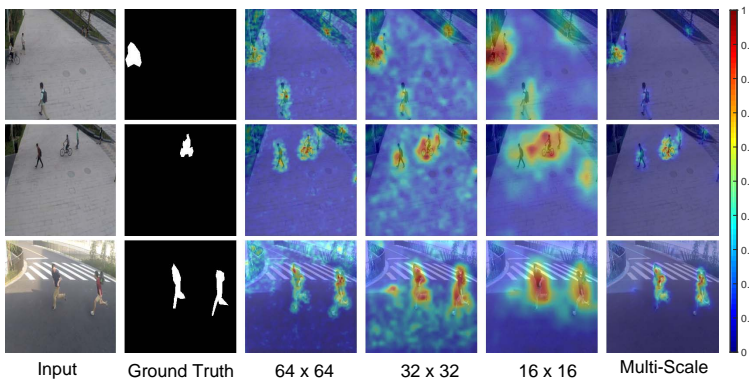


Figure 1: Visual results of our method on three anomalous images from the STC dataset. ResNet-18 is used as the backbone and the three bottom blocks (*i.e.*, conv2\_x, conv3\_x, conv4\_x) are selected as feature extractors. Columns from left to right correspond to input images, ground truth regions, anomaly maps of the three blocks, and the combined anomaly maps respectively.

Finally, we test whether our method generalizes well on small-sized images. We conduct experiments on CIFAR-10 which is traditionally used for image-level one-class classification. For this dataset, the sizes of feature maps from three blocks are  $8 \times 8$ ,  $4 \times 4$ , and  $2 \times 2$  respectively. The hyper-parameter settings are the same as those used in the MVTec AD

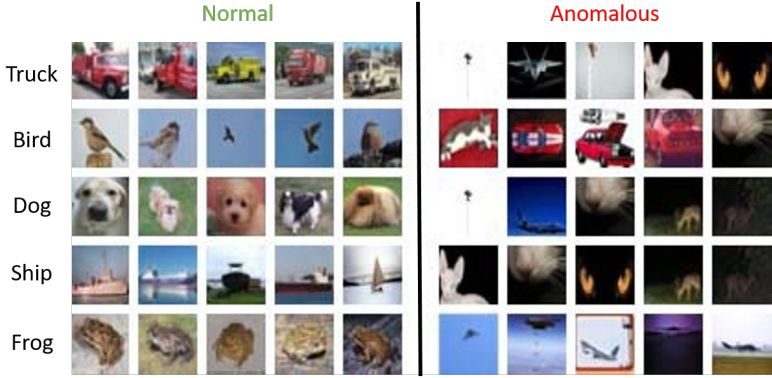


Figure 2: Results of our method on the CIFAR-10 dataset. ResNet-18 is used as the backbone and the three bottom blocks (*i.e.*, conv2\_x, conv3\_x, conv4\_x) are selected as feature extractors.

experiment. Following [1], we take one class as normal and the remaining nine classes as anomalous in turn, totally resulting in ten runs. We see clearly from Table 2 that our performance is much better. For each category, we rank the test images according to their anomaly scores. The top five samples and the bottom five sample for five categories are given in Figure 2. It is clear that our method predicts the images belonging to the given category as normal and the ones from the other categories are correctly classified as anomalies.

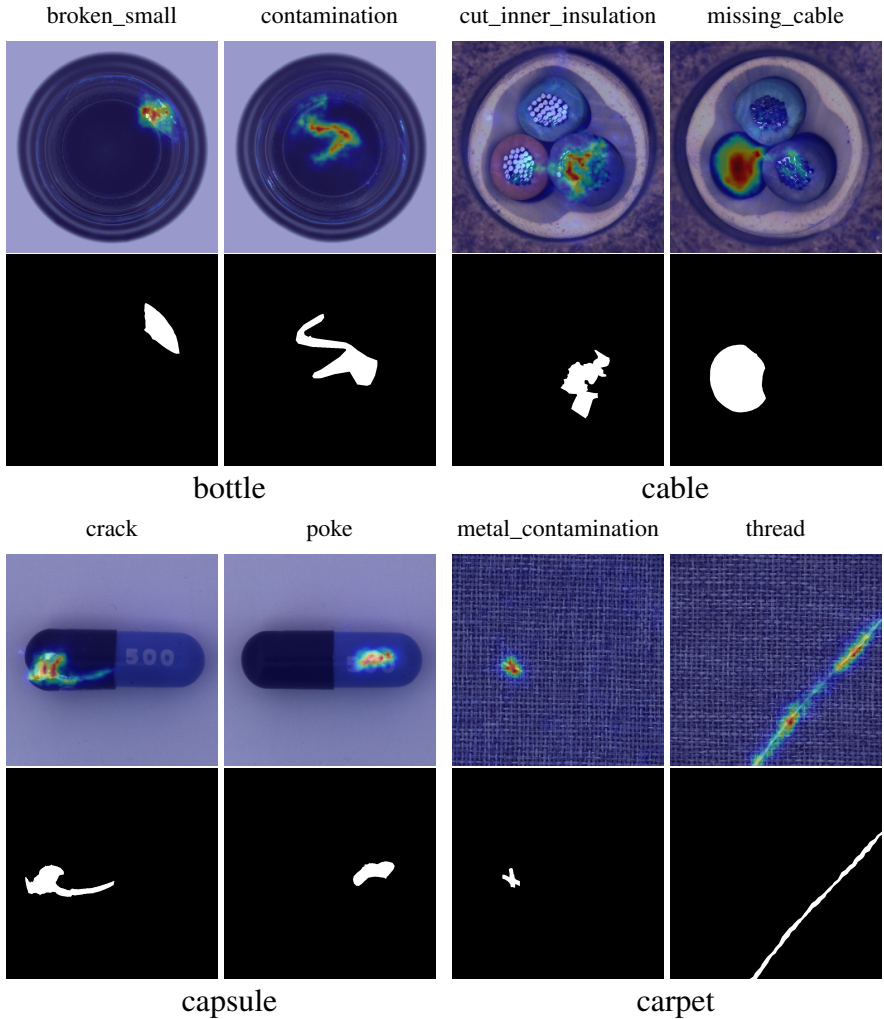


Figure 3: More visual results of our method on some example images from the MVTec AD dataset. Higher confidence in anomalous pixels are displayed in redder color. Zoomed in for better display.

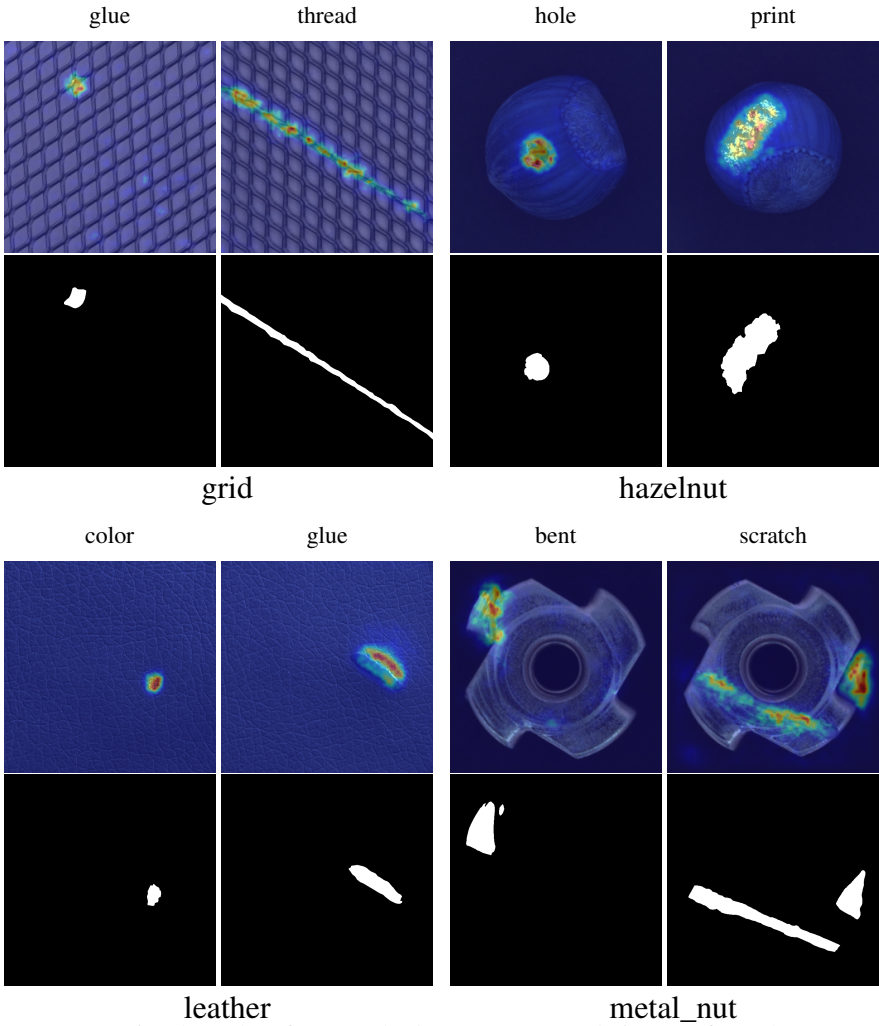


Figure 4: More visual results of our method on some example images from the MVTec AD dataset. Higher confidence in anomalous pixels are displayed in redder color. Zoomed in for better display.

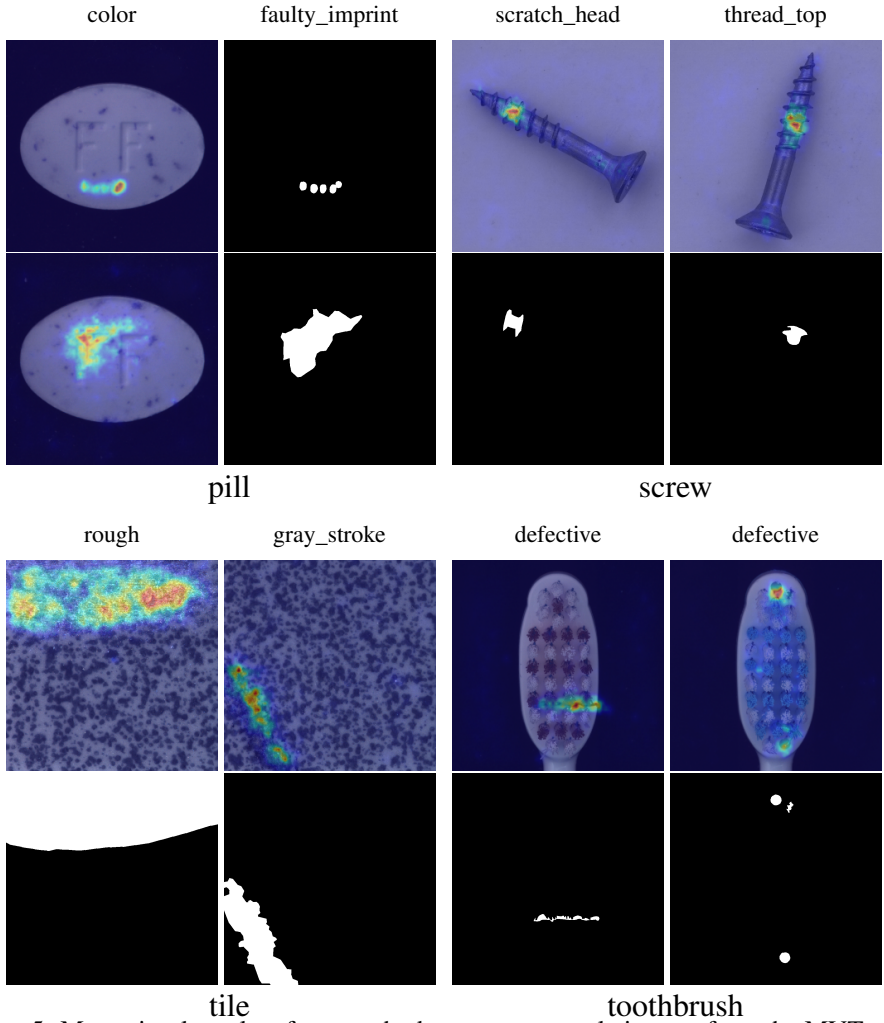


Figure 5: More visual results of our method on some example images from the MVTeC AD dataset. Higher confidence in anomalous pixels are displayed in redder color. Zoomed in for better display.

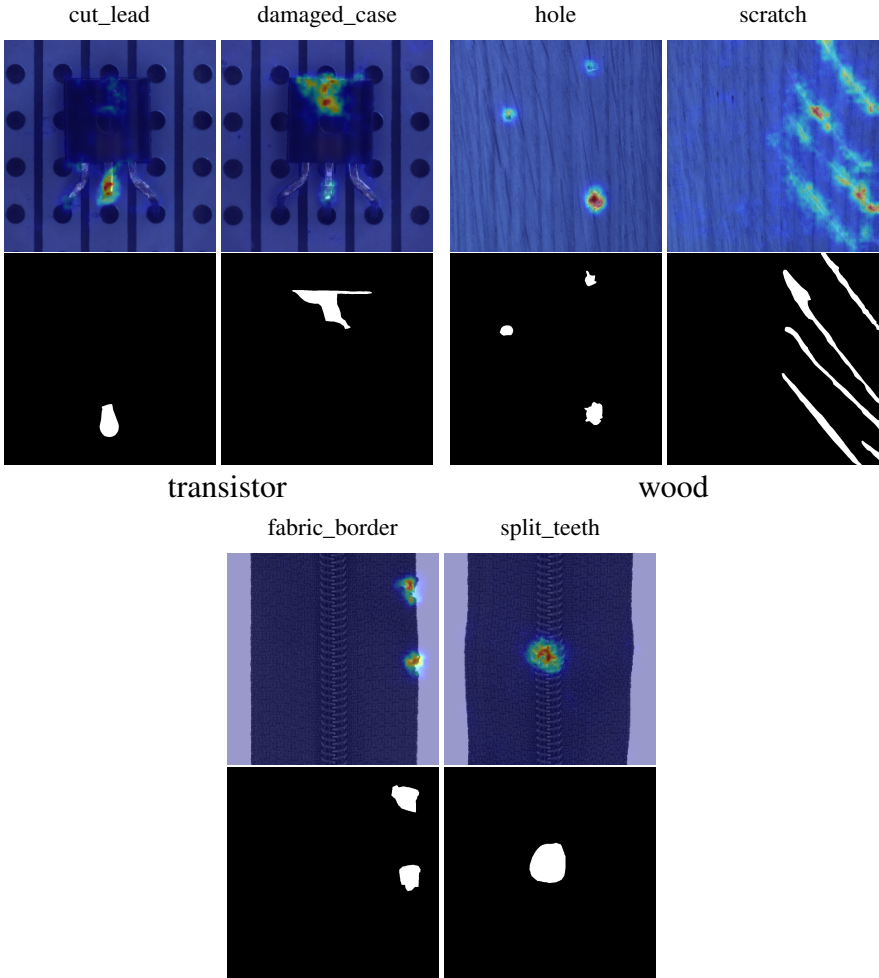


Figure 6: More visual results of our method on some example images from the MVTec AD dataset. Higher confidence in anomalous pixels are displayed in redder color. Zoomed in for better display.

## References

- [1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, 2019.
- [2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *CVPR*, 2020.
- [3] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv:2005.02357*, 2020.
- [4] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.
- [5] Weixin Luo, Wen Liu, and Shenghua Gao. A revisit of sparse coding based anomaly detection in stacked rnn framework. In *ICCV*, 2017.
- [6] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *ECCV*, 2020.